Università degli Studi di Pavia

Dipartimento di Studi Umanistici

Corso di Laurea Magistrale in Linguistica Teorica, Applicata e delle Lingue Moderne

ENGLISH-LANGUAGE CYBERDUBBING IN THE AGE OF ARTIFICIAL INTELLIGENCE: A RECEPTION STUDY

RELATORE

Prof.ssa Maria Gabriella Pavesi

CORRELATORE

Prof.ssa Claudia Roberta Combei

Tesi di Laurea Magistrale di

Lorenzo Costabile

Matricola n. 522487

Anno accademico 2023/2024

INTRODUCTION 6

CHAPTER 1 AUDIOVISUAL TRANSLATION FROM PAST TO PRESENT 9

1.1. AUDIOVISUAL TRANSLATION 9

1.1.1. MULTIMODAL AND MULTIMEDIAL TEXTS 10

1.1.2. AUDIOVISUAL TRANSLATION STUDIES 12

1.2. AUDIOVISUAL TRANSLATION TECHNIQUES: AN OVERVIEW 13

1.2.1. SUBTITLING 14

1.2.1.1. RESPEAKING 21

1.2.2. REVOICING 22

1.2.2.1. LIP SYNCHRONISED DUBBING 22

1.2.2.2. VOICE-OVER 22

1.2.2.3. SIMULTANEOUS INTERPRETING, FREE COMMENTARY AND NARRATION 24

1.2.2.4. AUDIO DESCRIPTION FOR ACCESSIBILITY PURPOSES 25

1.3. DUBBING 27

1.3.1. EVOLUTION OF DUBBING TO THE PRESENT DAY 27

1.3.2. DUBBING IS NOT DEAD: THE REBIRTH OF ENGLISH-LANGUAGE DUBBING. 32

1.3.3. TECHNICAL PROCESS AND DIFFICULTIES 35

1.3.4. FEATURES OF DUBBED LANGUAGE 37

1.3.5. DUBBING VS. SUBTITLING: A NEVER-ENDING STORY? 44

CHAPTER 2 AUDIOVISUAL TRANSLATION IN THE AGE OF ARTIFICIAL INTELLIGENCE 47

2.1. ARTIFICIAL INTELLIGENCE REVOLUTION 47

2.1.1. ISSUES AND RISKS 49

2.2. AUTOMATIC DUBBING 53

2.2.1. UNDERSTANDING SYNTHESISED SPEECH 55

2.3. NEW CYBERDUBBING CULTURES 57

2.4. RESEARCH QUESTIONS 65

CHAPTER 3 A RECEPTION STUDY 67

3.1. STUDYING AUDIENCES 67

3.1.1. PERCEPTION VS. RECEPTION 67

3.1.2. RECEPTION THEORY 68

3.1.3. RECEPTION AND AUDIOVISUAL TRANSLATION 70

3.2. MATERIALS AND METHODS 73

3.2.1. VIDEO SELECTION 73

3.2.1.1. CLIP 1 78
3.2.1.2. CLIP 2 78
3.2.1.3. CLIP 3 79
3.2.1.4. CLIP 4 79
3.2.2. METHODOLOGY 79
3.2.2.1. PARTICIPANTS IN THE RECEPTION STUDY: DEMOGRAPHICS AND VIEWING HABITS 81
3.2.2.2. PROCEDURE OF THE RECEPTION STUDY 84
3.2.2.3. DATA ANALYSIS 86

CHAPTER 4 RESULTS AND DISCUSSION 89

4.1. PERCEIVED NATURALNESS 89
4.1.1. ACCENT 90
4.1.2. AUDIO QUALITY 92
4.1.3. LANGUAGE 93
4.1.4. OTHER CONCEPTUALISATIONS OF NATURALNESS 95
4.2. SYNCHRONISATION AND VOICE MATCHING 97
4.2.1. LIP SYNCHRONISATION 97
4.2.2. SPEECH-TO-BODY CORRESPONDENCE 100
4.3. EMOTIONAL TONE 103
4.4. ATTITUDES AND OPINIONS 106
4.4.1. IDENTIFICATION TASK 106
4.4.2. CONCLUDING REMARKS ON THE RECEPTION STUDY 108
4.4.3. WHAT FUTURES FOR AUDIOVISUAL TRANSLATION? 109

CONCLUSIONS 115

APPENDICES 120

APPENDIX A: INFORMATIVE QUESTIONNAIRE 120
APPENDIX B: INTERVIEW SCHEME 121
APPENDIX C: INFORMED CONSENT 122

REFERENCES 123

SITOGRAPHY 133

FILMOGRAPHY 137

INTRODUCTION

The contemporary multimedia landscape is more diverse and fragmented than ever before. In the modern globalised and interconnected world, the ways of consuming audiovisual content are multiple and vary from individual to individual, thanks to widespread Internet access and technological developments. New fee-based streaming platforms, for example, have caused exponential growth in both content production and demand. This has significant implications for the audiovisual translation industry, given the rapid pace at which audiovisual content is released, which makes it challenging to translate/localise such material for distribution in diverse global markets. On their side, of course, industry players do not want to lose potential earnings from specific audience groups left uncovered. One consequence of this effort is the (re)emergence of English-language dubbing, which represents a significant shift from the traditional approach to translating foreign films and TV series in Anglophone countries, where subtitling is traditionally the preferred method. The streaming giant Netflix, for instance, has invested heavily in English-language dubbing with the objective of disseminating content from foreign (e.g. South Korea, Brazil, Spain) to English-speaking countries, and maximizing its reach. On parallel, social networks caused an analogous (and even greater) increase in the production and consume of audiovisual material. A new phenomenon that is emerging in this complex environment is the practice of cyberdubbing. While cybersubtitling (the creation of subtitles by online users for other online users) has often been the subject of study in the field of audiovisual translation studies, cyberdubbing represents a smaller and mostly unexplored niche. Cyberdubbbing is experiencing a significant surge in growth, particularly due to the advent of novel technologies. In particular, artificial intelligence is currently affecting a wide range of domains, including that of audiovisual translation. Today, the incorporation of AI-based tools into the professional processes of AVT seems to be unstoppable (since it allows to accelerate tasks while reducing costs). This opens up new lines of research in AVT studies for the analysis of the implications of this technology in the professional sphere as well as in the user side.

In light of these observation, the present study was elaborated. As a matter of fact, due to the novelty of the three abovementioned phenomena (i.e. English-language dubbing, AI dubbing, and cyberdubbing), it appears that there is a gap in the existing literature on this topic. To examine aspects within the field of AVT, reception studies are helpful as they provide authentic data from actual audiences. In this context, audience reception was identified as the appropriate strand of research to investigate new types of dubbing. Thus, an attempt will be made to study, through qualitative analysis methods, the audience's reception of (English-language) cyberdubbing generated with AI-tools.

Prior to undertaking the actual reception study, the thesis will provide a comprehensive overview of the world of audiovisual translation (AVT). Indeed, it would be impossible to engage in a discourse on AVT without understanding its historical development and defining characteristics. Chapter 1, thus, will be entirely dedicated to the systematic description of the various AVT techniques. Following an explanation of the term 'audiovisual translation' and its associated field of study, the subsequent sections will analyse two major areas: subtitling and revoicing. The history of subtitling spans from the earliest days of cinema history to the present, where technologies have enabled automatic subtitling. This section will address different themes, including constraints inherent to this AVT method, accessibility issues, and subtitling's potential applications, particularly in the field of language learning. The second section is that of revoicing, a macro-category which encompasses a range of techniques, including audio description (for accessibility purposes), free commentary, narration, simultaneous interpreting, voice-over, and, of course, lip synchronised dubbing. An entire section will be then devoted to dubbing, as it represents the subject of the study. The dubbing section will dwell on the historical development of the method, from its origins (coinciding with the advent of sound in film) to its contemporary manifestations, with a particular focus on the novel phenomenon of English-language dubbing. Since it is very often considered as a sort of 'constrained' translation, all the limitations and problems that dubbing poses will be explored, as well as the characteristic features of the dubbed language.

The second chapter will attempt to provide an exhaustive overview of the developments of artificial intelligence (AI) in the field of AVT. First, it will be necessary to introduce AI in general, and its associated issues and risks. Subsequently, special emphasis will be placed on automatic dubbing. On the one hand, the changes that this technology introduces in the AVT professional industry will be explored; on the other, the opportunities offered by state-of-the-art automatic software for dubbing will also be mentioned. The technology's workflow will be explained, as well as the difference between this and general text-to-speech workflows. To enrich the discussion, examples of studies and projects aimed at developing or perfectioning the speech synthesis technology will be provided. After that, a section will be devoted to new cyberdubbing cultures. Specifically, a survey on social networks will be conducted in order to identify content creators engaged in the production of automatic dubbings from Italian into English. Following the selection of some specific profiles on the social network Instagram, an analysis of their content will be carried out, along with an examination of the production processes involved. Indeed, an informative questionnaire will be proposed to the admins of these Instagram pages, to collect data that could add to the discourse on cyberdubbing practices. Afterwards, the research questions will be listed. As it will be seen, all the research questions are intended to shed light on aspects deemed particularly impactful for the contemporary and near future in the field of AVT.

Chapter 3 will be dedicated to a description of the materials and methods used for the study. Initially, a thorough examination of the concept of 'reception' will be provided, along with examples of relevant reception studies. The 'materials' section will consist of an explanation of the videos used for the reception study, i.e. two automatically dubbed videos selected from one of the Instagram profiles previously identified ('italiancomedydub') and two other excerpts from Italian films dubbed into English by professionals. Finally, the methodology will be addressed. As it will be seen, to assess the quality of the selected clips and the overall opinions and attitudes towards the use of AI in AVT, a qualitative semi-structured interview will be proposed. First, the sample of the participants to the reception study will be defined. Then, the procedure of the interview will be illustrated step-by-step: after the viewing of the excerpts, pre-determined rounds of open questions will be posed to the respondents, in order to look both at texts and beyond texts. Although results from this type of qualitative research method are not generalisable on a large scale, the interview will be used as it encourages detailed responses from the audience, providing elaborate and detailed data that can be confirmed in potential further quantitative research.

The fourth and final chapter will be concerned with the discussion of the results. The analysis will be based on the 'thematic analysis' approach, which consists of identifying key themes arising from trends and patterns within the comments elicited in the interview.

Overall, considering the scarce research in this still young area, the study hopes to intercept audience preferences in AVT and to shed light on the current (and future) state of AI-dubbing technology. Despite the limitations that will be described later, the results will provide an uncommon angle on the latest trends in AVT.

CHAPTER 1
AUDIOVISUAL TRANSLATION FROM PAST TO PRESENT

1.1. Audiovisual Translation

Recent decades have witnessed a global and unprecedented proliferation of audiovisual (AV) products. Films and TV series that the public loves to binge-watch online, video games that millions of people both play and watch others play via streaming platforms, long-format video podcasts on YouTube, as well as the endless flow of content on social-media apps are merely a small fraction of a potentially limitless list. The reasons behind this phenomenon are many and will be discussed later. What should be emphasised now is that the increase of AV products' consumption is paralleled by a spread of audiovisual translation (AVT) (Chaume 2018: 41). Before focusing on the aims and on the theoretical framework of the present study, it is necessary to pose some preliminary questions, such as 'What is audiovisual translation?' and 'How is it different from other types of translation?'. The study is, as a matter of fact, interested in tackling problems and new trends of AVT. For this reason, it would be impossible to begin without a proper description of what audiovisual translation consists of, how it works, what are its different areas of interest, and why it is important to study it at the present time.

AVT is an ever-evolving area, and this is also the reason why clear-cut definitions and labels are often difficult to trace within it. Chaume (2013: 106) provides a list of possible terms that have been used by different scholars to refer to audiovisual translation throughout its history, including film translation, media translation, multimedia translation and transadaptation, among others. As Pérez-González summarises, AVT is a type of translation "concerned with the transfer of multimodal and multimedial texts into another language and/or culture" (2019: 30). Let us now see why AVT can be regarded as radically different from other types of translation.

In 1959, Russian-American linguist and semiotician Roman Jakobson published an essay titled *On Linguistic Aspects of Translation*, which can be used as a starting point to examine translation in an all-encompassing way. In this seminal study, Jakobson distinguishes three possible ways to interpret a linguistic verbal sign, which is to say three different types of translation. The first type is the intralingual translation, in which a sign from a given language is interpreted and replaced by another one belonging to the same linguistic system. This type of rewording happens, for instance, when we reformulate words according to the level of formality required in a communicative situation. The second type of translation is interlingual, the so-called "translation proper". Indeed, this involves the replacing of signs from one language with others from a different language. Thirdly, Jakobson mentions the intersemiotic translation (or transmutation), which consists in the "interpretation of verbal signs by means of signs of nonverbal sign system" (Jakobson 1959: 233). This last type of translation goes beyond the mere word-to-word equivalence and aims at conveying

overall meaning. A traditional case is the book-to-film translation. However, Jakobson did not directly exemplified intersemiotic translation in this or later works. On the other hand, academic interest on intersemiotic translation (often referred to as 'adaptation') today has increased significantly. This "is certainly due to some market trends of the last decades, when […] the entertainment industry has been investing heavily in the production of different versions of already existing works for the most diverse audiences" (Da Silva 2017: 74). In this context, adaptations, reboots of classic films or alternative national versions of foreign successes are all examples that highlight the importance of the translation processes in the new media landscape. "Jakobson developed most of his work between the 1920s and 1960s. At that time, the use of images for human communication, in the way we understand currently, was still limited" (2017: 80). Today, a vast array of text types in which several sign systems coexist (like TV series, video games and films) has emerged, questioning the traditional tripartition proposed by Jakobson. We will later see how AVT can be either intralingual, interlingual or intersemiotic, depending on the translation process and the types of multimodal elements included. It is in this context that audiovisual translation emerges as a fundamentally new, multifaceted and dynamic branch of translation studies. Unlike the written texts *tout court*, audiovisual texts are in effect multifaceted by nature and require, for this reason, an in-depth analysis.

1.1.1. Multimodal and multimedial texts

In order to fully comprehend the complexity of AVT, it is necessary to examine the objects it deals with. Following the definition of AVT provided by Pérez-González above, a first distinction between the terms 'multimodality' and 'multimediality' can be useful to understand the types of texts that are addressed by this type of translation. As already mentioned, the interest in academia towards AV products has grown exponentially in the last decades, and many scholars have contributed to defining characteristics of these texts, and how they build meaning.

Kress and Van Leeuwen define multimodality as "the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined" (2001: 20). This description does not tell us much, however, if we do not specify what is meant by the words 'semiotic modes'. In another text about multimodality, Kress describes a mode as a "socially shaped and culturally given semiotic resource for making meaning", adding that "image, writing, layout, music, gesture, speech, moving image, soundtrack and 3D objects are [all] examples of modes used in representation and communication" (2010: 79). According to Stöckl (2004), in order to organise modes in a hierarchical structure, it would be possible to identify four core modes, namely language, sound, music and image, around which several sub-modes are connected. Core modes are, essentially, those "resources that we intuitively fall back on to articulate

our opinions on the audiovisual texts that we consume or produce. Unsurprisingly, the semiotic contribution of core modes to the overall message conveyed by films tends to feature prominently in most reviews" (Pérez-González 2014: 192). In *Ennio* (Tornatore 2020), the documentary film about the legendary Italian composer Ennio Morricone, there is a passage of an interview in which Morricone himself offers an insight into the importance of his creative and decision-making progress for the success of Elio Petri's *Indagine su un cittadino al di sopra di ogni sospetto* (1970):

> "Per *Indagine* io ho scritto un piccolo arpeggio. Durante il missaggio Elio mi chiamò e mi fece vedere tutta la prima parte, e io dissi: 'Ma, Elio, che è 'sta roba?'. Aveva cambiato la musica. Aveva messo la musica di un film orrendo che io avevo fatto anni prima. Elio mi diceva: 'È fantastica! […] Guarda come funziona 'sto coro' […]. E io dicevo: 'Ma possibile che vi piace 'sta cosa su 'sta scena? Ma che c'entra?'. A poco a poco c'è stata la mia resa assoluta e io ho detto: 'Fai come ti pare'. Si accese la luce e lui mi disse una cosa che io ricordo sempre con una certa commozione: 'Tu hai fatto la migliore musica che potevo immaginare, mi dovresti prendere a schiaffi'".

The extract well emphasises the role of music (as a core semiotic mode) in the construction of the overall meaning of a film – as, ultimately, Petri retained the music initially composed by the Maestro, and the film proved to be a success.

If all social and cultural objects and processes can be counted as meaning-making resources, it is true that multimodality has always existed, and that everything is to some extent multimodal. In a traditional printed book, for instance, the meaning is not just realised by the written words, but also by many semiotic modes such as "its typography (type and size of characters), its margins, its illustrated cover, the use of colours, and the presence of images, to name a few" (Gambier 2023: 2). Even in a simple spoken interaction, verbal signs are accompanied by gestures, facial expressions, and other meaning-making resources. Nevertheless, it is also undeniable that, more than ever before, the modern world relies on texts that are multimodal by definition (like films or websites) for entertainment or communication purposes. The question thus arises to what is actually meant by the term 'multimodal text'. Firstly, it is vital to remember that "texts are not limited to the spoken and written media of language. Instead there are many other resources that can be used to create texts in addition to the spoken and written word" (Baldry and Thibault 2006: 4). Multimodal texts are, thus, texts that display properties of multimodality. Audiovisual texts, in particular, are multimodal specifically because their very production and interpretation is tied to several semiotic modes. Therefore, "for specialists coming to translation studies with a background in the study of modern languages, analysing meaning that is conveyed partially or totally through non-verbal semiotics is particularly challenging" (Pérez-González 2014: 142).

The other term previously mentioned is 'multimediality', which should not be considered as a synonymous of 'multimodality'. Let us see at two clarificatory examples provided by Kress and Van Leeuwen: "radio […] is multimodal in its affordances, because it involves speech, music and other sounds; but it is monomedial, since it can only be heard, and not seen, smelled, touched or tasted. Everyday face to face interaction, on the other hand, is both multimodal […] and multimedial (it addresses the eye and the ear and potentially also touch, smell and taste)" (2001: 67). Therefore, it is clear how "audiovisual texts are multimedial in so far as this panoply of semiotic modes is delivered to the viewer through various media in a synchronized manner, with the screen playing a coordinating role in the presentation process" (Pérez-González 2019: 30).

AVT experts work with texts that are both multimodal and multimedial, and this is exactly why AVT is dynamic, multifaceted, and different from other types of translation.

1.1.2. Audiovisual translation studies

AVT is a sub-discipline of the wider discipline of translation studies (TS). On the one hand, it constitutes a separate subtopic as it specifically deals with audiovisual products only. On the other hand, in research terms, AVT is highly intersectional as it draws from a wide range of disciplines to develop its theories, including, for instance, sociolinguistics, pragmatics, psychology, accessibility studies, and reception studies. Thanks to this broad multidisciplinary base, "AVT has been able to diversify its methodological toolkit, an essential step in addressing the challenges […] of the massive growth in the volume of cultural artefacts mediated via AVT in both mainstream and amateur practices" (Guillot 2020: 318). Indeed, the area has gained a prominent role in TS due to the unprecedented spread in the demand (and production) of AV products, which are no longer restricted to films and TV shows.

Among the many scholars who have addressed the reasons behind the explosion of AVT, Gambier (2023: 1) states that technology (and its undisputed importance for the modern and globalized world) is currently challenging the traditional Western concept of translation. On an analogous note, Pérez-González (2014) also mentions digital technology development as the main contributor for the establishment of AVT as a standalone discipline:

> "the mutually shaping relationship between audiovisual translation and technological innovation has created a need for (i) robust theoretical frameworks to assist with the conceptualization of new text-types; (ii) new methodological approaches to guide the researcher through issues pertaining to the compilation, manipulation and analysis of samples of audiovisual data; and (iii) a better understanding of the new discourse communities formed around the production and consumption of established and emerging audiovisual text-types" (2014: 13).

Although the first publication to focus on AVT dates back to 1960 with a special issue of the journal *Babel*, the discipline is relatively young overall, as its inception is usually placed in the 1990s (Díaz-Cintas 2019: 215). Despite being recent, however, AVT now matches the status of other areas of study within the broader field of TS (Chaume 2018: 41). A detailed account of the development of AVT research is provided by Chaume (2018), who explains how there have been four main methodological paradigms to date.

The first stage was the Descriptive Translation Studies (DTS), which, from the late 1990s and early 2000s onwards, began to research translation norms, strategies and methods in AVT, in the same way as it had been done for traditional literary translation in the past. Analysing dubbings and subtitling operation through the use of corpora, DTS sought objectivity in the micro- and representativeness for the macro-dimension. In the 2010s there was the Cultural Turn, i.e. aspects beyond the actual translation operations began to be investigated, such as the ideological dimension in dictatorships. The third turn has been that of Sociological Studies, which aimed at analysing the translator's role, the working process in specific settings, the market and its fluxes, as well as the audience and its habits of consumption (using questionnaires, interviews or focus groups). Recently, thanks to technological advancements, one last turn has occurred. It is the case of Cognitive Studies in AVT, which objective is to explore cognitive processes of translators and audiences through new devices such as eye-trackers and other biometric sensors (2018: 54).

All these methodological paradigms can be used today to investigate phenomena within the AVT landscape, which continues to evolve as the production and consumption of screen-based texts is more active than ever before. Since the phenomenon has only recently emerged, for instance, there is still scarce production in academic literature on the relationship between AVT and artificial intelligence (AI). This will be attempted in the present study, whose methodology will be delineated in detail in Chapter 3.

1.2. Audiovisual translation techniques: an overview

Once established the peculiarities of the discipline, it seems now appropriate to discuss about the actual techniques that constitute AVT itself, whose complexity is evident as it branches out in many different directions. As a matter of fact, the term AVT is not strictly referred to a single procedure, but rather to a heterogeneous ensemble of techniques. While subtitling and dubbing remain the most important ones as of today, the vastness of this diverse field requires a more extensive analysis. The complex set of transfer methods can be organized under two principal headings: subtitling and revoicing. The two categories incorporate a number of methods that will be outlined later, including those for accessibility – a theme of paramount importance in today's world,

especially in online domains where user-generated content often raises accessibility issues. In the following sections, we will primarily present such techniques from a procedural perspective, highlighting their scopes as well as the differences existing between them. In addition to this, we will provide some historical background of the practices. To gain an overall understanding of the subject, in effect, it may be useful to collocate diachronically the various steps that ultimately led into the emergence of this branch of translation as it appears today.

1.2.1. Subtitling

Subtitling consists of the rendering of spoken dialogues into written text through snippets that are displayed on screen concurrently with the corresponding line of the spoken dialogue. In its various realisations, together with dubbing, it is today one of the two most widely used AVT practices globally. Its history, however, extends well beyond recent times, dating back a century to its earliest applications in cinema.

Intertitles of silent films in the 1920s can be considered as the ancestors of modern subtitles (Pérez-González 2020: 31) and as the first form of AVT (Chaume 2012: 11). It is for this reason that an examination of the history of subtitles can be useful to understand the evolution of AVT as a whole. As highlighted by O'Sullivan and Cornu (2019), "in the pre-sound era, films were silent, but not speechless" (2019: 15). Indeed, at the beginning of the 20th century, intertitles were inserted between film frames to facilitate the narration by providing information about the characters and the plot (Kapsaskis 2020: 554). Although silent, exporting these films to foreign markets still required some form of interlingual mediation. This process was obtained by simply removing the original intertitles and inserting new sets in the target language back into the film (Pérez-González 2020: 30). These text fragments, also known as title cards, were short sentences hand-drawn on black cards, subsequently filmed and inserted in between different scenes. When the sound film market began to expand in the US, the issue of exporting the so-called 'talkies'[1] to Europe was initially solved using intertitles as interlingual translation tools. The period in question saw the advent of what were known as 'synchronised films', silent versions of talkies with music as their only sound and inserted intertitles in the target languages (O'Sullivan and Cornu 2019: 17). This phenomenon was due to technical reasons (the dubbing process had not yet been developed) and, in some countries, also to political decisions. For instance, in Italy, a fascist law in 1930 banned all films containing speech in foreign languages, "sia pure in misura minima" (Quargnolo 2000: 19). Thus, the dialogues of the original films were muted and replaced by silent intertitles in between scenes. On the other hand, in spite of the advent of dubbing, subtitling continued to be the

[1] Talkies, or talking movies, were sound films which included synchronized dialogue, made during the period when most films were silent.

preferred AVT mode in many countries, mainly for its cost-effectiveness (Kapsaskis 2020: 554). Díaz-Cintas describes subtitling as "fast, inexpensive, flexible and easy to produce, […] the perfect translation ally of globalization and the preferred mode of AVT on the world wide web" (2012: 288). Indeed, the field of subtitling witnessed rapid advancement throughout the latter half of the 20th century and the first decades of the 21st century. "Today a standard PC with subtitling software is all that is needed for subtitlers to complete a film subtitling project. […] We are a long way away from subtitles' post-silent films intertitles debuts, […], and a long way from the timecoded VHS tapes of the 70s and 80s" (Guillot 2019: 32). AVT, however, developed differently in different countries.

Subtitling did not take root at the same way in all parts of the world, and even just in Europe the situation was and still is very fragmented. The main reasons are certainly economic, as countries with more financial possibilities and major film industries, such as Germany, Italy and France, have always tended to favour dubbing. In the case of bilingual communities (Belgium and the Netherlands) and smaller countries with smaller film markets (Scandinavian countries, Portugal, Greece, Iran and most Arab countries), subtitling has always been preferred, and the costly method of dubbing has never been adopted in an extensive way, at least for what concerns films' release in theatres (Pérez-González 2009: 18). In addition to economic factors, the choice of an AVT technique over another is also influenced by complex cultural and political reasons (Perego et al. 2018): similarly to what happened in Italy, Spain faced a dictatorship that banned the showing of foreign films in their original language. While sometimes problematic with subtitles, dubbing was useful to easily manipulate dialogues that had to be censored (Zabalbeascoa et al. 2001).

Nevertheless, subtitling market has been gaining attractiveness in the last few decades, even in the so-called 'dubbing countries' (see, for example, Chaume 2013, Matamala et al. 2017). Today, as Ghia and Pavesi (2021) point out with the support of a wide range of studies, there is a notable worldwide shift towards the use of subtitles. This phenomenon is especially true for film connoisseurs and younger demographics, but it is gradually expanding to encompass a wider range of audience. One key factor contributing to this shift is the proliferation of streaming platforms, where subtitled versions often become available much earlier than their dubbed counterparts. Contemporary trends in the distribution of AVT products are not, however, the only explanations for the increasing subtitles' preference by global audiences, as several other reasons could be investigated. As a matter of fact, new tendencies have emerged in the production and the consumption of AVT products as well. Much research conducted throughout the past few decades within the field of AVT (e.g. Pérez-González 2007, Díaz-Cintas and Muñoz Sánchez 2006) has concentrated, for instance, on the phenomenon of funsubbing, "subtitling by the people for the people" (Guillot 2019: 31). Reshaping the role of subtitling also in traditionally dubbing country like Italy (Massida and Casarini 2017), fansubbing is important to mention as a significant part of

AVT's recent and contemporary history. Emerged in the US in the 1980s, funsubbing culture developed as a reaction against the linguistic and cultural neutralization of Japanese anime (2019: 37). Starting from here, an ever-increasing number of networked amateur communities has established on online-grounded channels. Even within this cyber-subculture, there are numerous branches. Some endeavour to provide their peers with quality subtitles for content that is unavailable in certain countries. Others engage in this practice for creative and entertaining purposes on social media. Additionally, there are groups who, as previously mentioned, seek to provide their fellow fans with more authentic subtitles that do not conform to the insensitivity of mainstream commercial subtitling conventions (Pérez-González 2014: 17, Bruti 2019: 201). As Baños and Díaz-Cintas (2023) notice, fansubbing has helped change habits in the industry:

> "in a direct response to the appealing fast turnaround of subtitles created by fansubbers, large corporations and distributors have shortened the timespan between the release of original audiovisual footage and its localised editions, which are now launched at the same time as the domestic product in a strategy called simultaneous shipping (simship), as well as diversified the options for accessing this content, giving birth to the nowadays popular binge watching experience" (2023: 138-139).

This discussion will be resumed later, as many aspects of this topic can be extended to the domain of dubbing, considering the emergence of fandubbing as a novel phenomenon.

Having outlined the historical development of subtitling, it is now appropriate to ask: what, exactly, does this technique entail? In order to analyse of the multiple facets of this AVT mode, we could start from the definition provided by Díaz-Cintas and Remael (2007) in their influential text about subtitling. Here, they define subtitling as:

> "a translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavours to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, inserts, graffiti, inscriptions, placards and the like) and the information that is contained on the soundtrack (songs, voices off)" (2007: 8).

The first element that warrants consideration is the clause "generally on the lower part of the screen". Indeed, subtitles are typically placed at the bottom-centre of the frame. However, in recent years there has been an increase in the use of experimental subtitles, positioned in non-traditional ways. For what concerns film subtitles, it is becoming increasingly common to see subtitles moving to the right or left of the screen according to the character speaking at that particular moment.

Another trend is related to recent changes in media viewing behaviour. As a matter of fact, the direct consequence of the increasing shift towards short-form video content on social media is the users' exposure to new types of subtitles, used extensively or the creation of memes in video format. Finally, with the spread of streaming platforms, users are also allowed to experiment freely and adapt subtitles to their preferences[2]. According to some studies (e.g. Crabb and Hanson 2016, Crabb et al. 2015), placing 'dynamic subtitles' (opposed to traditional ones) in varying positions according to personal preferences and different video contexts can be positively impactful for a more immersive and less disruptive viewing experience. Berke et al. (2019) point out that readability and occlusion are the two key concepts to consider in the context of subtitles customisation. The main problem consists in the existing tension between these two concepts, since, for example, a black box may help the reading of lines but obstruct the visibility of significant portions of video content. The relation between readability and occlusion is echoed and amplified in many different media landscapes, outside the traditional ones. Rothe et al. (2018) mention the viewing experience in the new field of cinematic virtual reality, where the solution to aid viewers' comprehension and prevent dizziness is thought to be 'worldreferenced subtitling', i.e. the positioning of subtitles in close proximity to the speaking person in the 360° movie environment. Regarding social networks, McDonnell et al. (2024) analyse the complex context of TikTok, the leading social network based on user-generated short videos. With the quick development of automatic speech recognition technologies, online platforms have adopted automatic closed captioning for video content[3]. Although TikTok has promptly implemented this option, the platform seems to display a culture of user-generated creative subtitling, with highly stylized open captions that catch viewers' attention. Given TikTok's popularity (2024: 3), this trend is reflected in other platforms as well, and has quickly led to an increased attention in the press towards Gen Z's 'incapability' of watching non-captioned content (e.g. Pogue 2024, Kelly 2023).

To prevent any ambiguity on the matter, it is best to make a distinction between subtitles and captions. Although the two terms are often employed interchangeably by non-experts, subtitles and captions are not quite synonyms. In fact, they have very distinct purposes. On one hand, basic subtitles' function is to make spoken dialogue visible, and thus, comprehensible to all viewers. On the other hand, captions include written descriptions of paralinguistic [1] and extralinguistic elements [2], as well as other relevant information within the audio track [3] used as narrative tools.

[1] *(INHALES DEEPLY)*, *(BURST OUT LAUGHING)*

[2] *(CAR APPROACHING)*, *(DOOR CREAKS OPEN)*

[2] American giant Amazon Prime Video, for instance, allows viewers to change colour and size of subtitles. YouTube, on the other hand, has almost completely customable subtitles, with options even for position and font type.

[3] A more detailed examination of this topic will be provided below, when discussing respeaking.

[3] *(OMINOUS MUSIC IN THE BACKGROUND), ("SONG TITLE" PLAYING)*

As mentioned above when talking about TikTok, captions can be either open or closed. Open captions are burned in the video, whereas closed captions are optional and can be switched on and off by viewers. Captions were first implemented as accessibility tools in 1950 by the association Captioned Films for the Deaf (Romero-Fresco 2020: 549). Today, streaming platforms such as Netflix include closed captioning (indicated by 'CC') in the subtitles' window to indicate subtitling intended for the deaf or hard-of-hearing (SDH). Captions' application, however, is no longer limited to accessibility for auditory impaired audiences. Referring to Jakobson's tripartition discussed in Section 1.1.1., captions can be regarded as a form of intralingual subtitles that are beneficial for a wider range of viewers. People with a lesser command of the language (2020: 549), such as immigrants in a foreign country or – more broadly – second language learners (SLLs), in effect, might need or want a transcription of the audio track. Unless further specified, the most traditional practice intended when talking about subtitling is interlingual subtitling (i.e. the insertion of the written translation of the source speech into a target language)[4]. However, intralingual subtitling is becoming increasingly popular, serving also as a means facilitate language learning (Neves 2019: 87).

Many are the studies that investigate the positive role of subtitles as learning tools in formal learning environments (e.g. Incalcaterra McLoughlin and Lertola 2011, Perego et al. 2015). In addition to this, it is proven that subtitles are useful and function well also in out-of-school contexts. This phenomenon is of particular interest in academic settings, and it is known by the name of 'informal language learning' (ILL). Most studies in this area refer to English informal learning. The reasons are many and are to be found in today's globalised and interconnected world. Ferguson (2015) describes the modern world as characterised by increased multilingualism and ethnic diversity. In this context, English gains ground and dominates the scenario as a global lingua franca, expanding its domains of application. SLLs also expand their learning domains, departing from the conventional boundaries of formal education and constructing what Benson defines as the "individual perspective" (2021: 7) of learning environment, namely a configuration of settings and resources assembled by the individual learner. As testified by Pavesi and Ghia (2020), in the past, the experiences of SLLs were primarily confined to conventional educational settings. Today, there has been a notable shift away from this, accompanied by the emergence of ILL as a prominent phenomenon. To be considered informal, this type of learning must occur outside of the school environment and be incidental, i.e. the learners' primary focus has to be on the message conveyed

[4] Another form of subtitling not covered here is bilingual subtitling, i.e. the displaying of subtitles in two different languages at the same time on screen, typically used in bilingual countries such as Belgium or Finland.

in the input, rather than on its form. Such exposure occurs, for instance, when one listens to music in a foreign language or watches subtitled films. As a matter of fact:

> "For some decades now, audiovisual input has been considered from a language learning perspective and is widely acknowledged as an effective resource for gaining literacy in an L2 also informally. […] When added to images, subtitles provide a useful support to general comprehension and contribute to lowering learners' affective filter […]. Concurrently, they promote the activation of matching processes between spoken dialogues, written text and images, whereby learner-viewers are stimulated to compare oral and written text and associate verbal input with visual elements, thus engaging in a multimodal and multisemiotic meaning-making process […]. The matching activity between semiotic resources leads to overall deeper processing of the input and can drive viewers' attention to form, potentially triggering noticing processes" (2020: 53).

In addition to this, the use of subtitles is often preferred also for reasons related to immersion, as they allow viewers to have a more authentic viewing experience. Although subtitles occupy a portion of the screen and can constitute an obstacle from this point of view, immersivity is mainly related here to the actors' performances. Indeed, according to Ghia and Pavesi, hearing the actors' original voices "and the whole palette of different shades and nuances of meaning" (2021: 173) constitutes an essential "hedonistic" aspect of film-watching. In the context of subtitling as an important aspect of the film and for the film, an insightful observation was put forth by Taylor's (2012). With regard to the broader discourse on multimodality, Taylor argues that subtitles, particularly same-language subtitles for learners or for individuals with auditory impairments, can be considered additive, i.e. an addition to an already multimodal product. This is because "while the addition of a piece of music or some background noise or a darkening of the screen would add something totally new, subtitles repropose, completely or in part, an already present semiotic modality" (2012: 26). To this regard, we could also mention the distinction proposed by Delabastita (1990: 102) between the concepts of 'adiectio' and 'substitutio', respectively associated with the subtitling and the dubbing technique.

There is an eternal debate (in academic and non-academic contexts) about the advantages and disadvantages of subtitling and dubbing (e.g. Díaz-Cintas 1999, Koolstra et al. 2002). On one hand, the "Cognitive Load Theory maintains that the redundancy effect obtained when information is presented simultaneously through more than one channel increases the cognitive load" (Incalcaterra McLoughlin 2019: 484). On the other hand, many studies argue that reading subtitles is a semiautomatic task which is not particularly demanding (Perego 2015: 2). Of course, these aspects vary according to a number of factors, including the quality of subtitles themselves

or the linguistic proficiency and cognitive capabilities of the viewer/reader. The debate is still an open one and will be addressed in greater detail later.

For what concerns the technical features of the modality, if it is true that subtitling is often preferred over dubbing due to its perceived more discreet nature, as it allows to enjoy the original acting performances in their entirety, it is also true that the original material does not remain untouched. Indeed, subtitling is a complex technique, and many are the challenges that it poses, the first of which are of technical nature.

In their comprehensive text on subtitling, Díaz-Cintas and Remael (2007) organise technical and formal issues under two major dimensions: the spatial and the temporal. As the authors underline, "subtitling is a type of translation that should not attract attention to itself" (2007: 82). Regarding the management of the limited screen space, it is important to reconsider here the tension existing between readability and visual occlusion. The number of lines – two lines at a time are preferred – is crucial, as too long texts can compromise the fruition experience and the comprehension of the message. Colour of lines is of equal importance: reading white characters on a bright white or grey background of a frame is often impossible, but directors do not make films thinking about how they will be subtitled. For such instances, the considerations on dynamic subtitles previously outlined remains applicable. In terms of temporality, succinctness is key: translators' work is constrained by time, because the subtitle projection must be aligned with both audio and video. First, the rhythm of subtitles has to reflect the rhythm of the recited dialogues, with strategies such as the splitting of long sentences across multiple subtitles and the combining of short sentences to prevent the pitfalls of telegraphic style (2007: 88). Another typical difficulty of the speech-to-writing transfer is when multiple voices overlap in the original audio track, and subtitlers need to decide what information needs to be written and what can be left out. Second, texts must match the underlying scene changes, as "poor timing, with subtitles that come in too early or too late […], detract from enjoying a programme" (2007: 90). Furthermore, physiological problems have to been taken into consideration. In effect, "audiences have limited time to read the text showcased and process it in its fragmented subtitle-by-subtitle sequential presentation" (Guillot 2020: 318), and this may result in the abovementioned supposed cognitive load. Psychological reception is also not a factor to be forgotten. As Marleau (1982) questions: "avec les apparitions et les disparitions brusques de quelque 900 sous-titres […], est-il possible d'apprécier un film à sa juste valeur?" (1982: 276). To minimise disruption, scholars and technicians established the 'six-second rule' as a recommended guideline. The rule states that viewers with no special needs should be able to read two full subtitle lines in six seconds. (Díaz-Cintas and Remael 2007: 96). Many of these technical issues of segmentation and synchronisation are assisted today by modern software. However, traditional translation problems persist and amplify in the subtitling process.

Of course, *mot-à-mot* translation must be avoided and texts, although dense and short, must be linguistically correct. In addition to this, since the original dialogues can be heard, it is particularly important (yet difficult) to respect all the linguistic characteristics chosen to represent the characters. As a matter of fact, "the way characters speak tells us something about their personality and background, through idiosyncrasies and through the socio-cultural and geographic markers in their speech, which affect grammar, syntax, lexicon, pronunciation, and intonation" (2007: 185). Finally, the translation of humour, quotations, cultural allusions and songs, when required for the purposes of the plot, represents a well-known challenging aspect. Pederson (2007) defines all these instances as 'translation crisis points', which confront translators with choices. The author focuses on what he calls 'ECR', i.e. extralinguistic culture-bound references, and on the various resolutive strategies that can be adopted. In this context, the case study presented by Anne Jäckel (2001) offers an interesting insight into the subtitling of the French film *La Haine*, by Mathieu Kassovitz (1995). The film well illustrates the difficulties of interlingual subtitling, as it presents all the issues abovementioned, as well as many others (such as the presence of swear words, which are problematic not only from the equivalence point of view, but also because their written form results more aggressive than their spoken counterpart). The language spoken by the protagonists (three young men of different ethnicities and characters, jobless and reliant on criminal expedients), is creatively and heavily marked, and reflect the cultural values of the reality the film represents, the difficult life in the *banlieue* of Paris. The film portrays, thus, "an almost perfect example of every possible deviation from standard French: sloppy language, bad grammar […], use of local colloquialisms, slang, verlan (backslang), Americanisms, Arabic, and all this intermingled with funk rhythm" (2001: 224). Under Kassovitz's instructions, the film was subtitled into English by cultural experts who adapted it almost entirely to American culture. The endeavours were, however, not well received by the audience. As Jäcke points out, many critics argued that all multiple meanings and nuances of the various speech forms had been lost in favour of a sloppy pastiche of Black American slang that prevented viewers from fully getting into the dimension of the film.

In summary, it can be stated that medium-related constraints mentioned before can, if not treated carefully, negatively affect the style, the rhythm, the personality, and ultimately the overall artistic and aesthetic outcome of the original work.

1.2.1.1. Respeaking

It is appropriate to conclude this overview of subtitling with a look at the present. Indeed, the modern technique of respeaking can also be classified as a form of subtitling. Respeaking, or real-time subtitling, consists of the live production of subtitles for audio files thanks to automatic speech recognition (ASR) technology, which entails the simultaneous conversion of waveforms into written texts. Similarly to general subtitling, respeaking is also an invaluable aid for individuals with

hearing impairment. However, if not carefully checked, it can present problematic errors, posing issues of accessibility. While traditional respeaking for the deaf and the hard of hearing is intralingual, over the past years, the process of interlingual respeaking has reached high performance levels as well (Romero-Fresco 2011: 12). Instances of such types of respeaking can be abundantly traced on social media like YouTube, where videos are accompanied by the so called 'auto-generated subtitles', created through machine learning algorithms. The specifics of the technique and the perceived risks associated with this and other modern AVT techniques will be discussed further in detail, in the section entirely dedicated to artificial intelligence and its use in the field of AVT today.

1.2.2. Revoicing

Revoicing is another major area of AVT whose importance cannot be underestimated. It should be mentioned that the label 'revoicing' is used, in fact, as a sort of umbrella term, since various different spoken translation methods can be located inside this category (Pérez-González 2014: 19), all based on the inserting of a new soundtrack to a pre-existing audiovisual product (Chaume 2013: 107). Analogously to the previous one, the present section provides an overview of such methods, considering their development as well as their technical features.

1.2.2.1. Lip synchronised dubbing

Technically speaking, lip synchronised dubbing (or, simply, dubbing), falls into the category of revoicing. However, we will not discuss this in detail in this section, for two main reasons. First, since dubbing (and its possible directions with the advent of AI) is the principal object of study treated in this work, a separate section will be dedicated to it. For its relevance in the film translation industry and its vastness as a field, dubbing is also often dealt separately in most academic texts about AVT. Second, dubbing is also considered almost as a distinct modality for technical reasons (Pérez-González 2020: 32). Indeed, although all revoicing methods involve some sort of synchronisation between on-screen images and audio tracks, dubbing is inextricably tied to this aspect. This need for synchronisation implies technical difficulties that other revoicing methods do not have. This aspect has led to the emergence of the label 'constrained translation' referred to dubbing (Pavesi 2020: 157), although several other issues related to the technique will be addressed later.

1.2.2.2. Voice-over

Voice-over is a type of revoicing that requires careful examination, especially in light of its difficult positioning within the field of AVT. From an historical perspective, it can be argued that the origins

of voice-over are the oldest in AVT. Indeed, the very first way to provide additional verbal language to films was introduced in the late 1890s and early 1900s, where the so-called 'film explainers' read title cards explaining the plot for illiterate audiences, providing cultural and technical details as well[5] (O'Sullivan and Cornu 2019: 16). Franco et al. (2010) explain how the performance of film explainers (also known as lecturers) can be considered, in essence, as a primitive voice-over. As many scholars point out (e.g. Díaz-Cintas and Orero 2006: 477), voice-over can be viewed from two different angles: that of translation studies (TS) and that of film studies (FS).

In the first case, voice-over is considered a proper transfer mode useful to adapt an audiovisual product from a source language (SL) to a target language (TL). The technique entails reading the translation of the SL dialogues in the TL while the original audio track, played at a reduced volume, is audible in the background. This technique could be seen, thus, as a sort of 'half-dubbing', since it is about overlaying the original audio track with a new one, instead of replacing it. Voice-over actors generally start reading the translation a few seconds after the start of the original audio track and finish a few seconds before its end. In this way, viewers are constantly aware that they are listening to a translation. This is often said to contribute to the feeling of authenticity and realism (Matamala 2019: 74). These features make it particularly useful for the translation of interviews, documentaries or other forms of AV texts where lip synchronisation is not essential. Even though a certain degree of synchronisation between sound and on-screen images is still required, independence from speech movements is an advantage that makes this AVT mode considerably cheaper and faster than traditional interlingual dubbing (2019: 68). Affordability is not a factor to be underestimated in the audiovisual industry. As a matter of fact, voice-over is the main AVT method in countries with small film markets in Eastern Europe (Poland, Romania), Middle East (Iran) or Asia (Thailand) (Pérez-González 2014: 20).

On the other hand, voice-over can be regarded as a form of narration that offers supplementary information about the ongoing audiovisual product. In this case, the technique can be situated within the broader field of FS, as no actual language transfer is carried out (Díaz-Cintas and Orero 2006: 477). This is not only the case of documentaries, since the technique is used in fiction in an analogous manner, with the function of guiding the audience towards an understanding of the film. Of course, third- and first-person narrations can be translated into other languages. Many categorise this type of revoicing under the name of 'narration', although more appropriate alternatives seem to be possible. The reasons behind this distinction will be explained in detail below.

[5] Film explainers read title cards of local films, provided translations of title cards of foreign films, and explained on-screen images. By doing all this, they performed intralinguistic, interlinguistic and even intersemiotic translation. These figures disappeared in the USA around 1920, but in countries like Japan they continued to exist until the 1930s. For the Japenese audience, film explainers (called 'benshi') were even more popular than the actual film stars.

1.2.2.3. Simultaneous interpreting, free commentary and narration

Three other types of revoicing often cause terminological confusion due to their similarities: it is the case of simultaneous interpreting, free commentary and narration. Let us now see why, although sharing some features, they are distinct in all other respects.

Simultaneous interpreting is a live revoicing method usually carried out in settings where more elaborate and expensive forms of revoicing are not possible. A typical situation is low-budget film festivals, where interpreters revoice live, with or without a script, the voices of the actors in a film or documentary (Pérez-González 2014: 20). Although similar to voice-over for the addition of a translation over an audible SL audio track in the background, there is no ambiguity in considering it a separate transfer mode. This is due to the fact that, as explicitly stated by its label, simultaneous interpreting is conducted on the spot (i.e. during the viewing event), whereas the voice-over procedure occurs separately and in a studio setting. Another fundamental difference is that voice-over involves (at least) two professionals: the translator and the voice-over actor. In contrast, simultaneous interpreting technique is carried out by the same person, the interpreter.

However, "a trickier relationship is that of voice-over with two additional transfer modes that are usually bundled together under the category 'revoicing', i.e. narration and (free) commentary, which share the absence of lip synchronization" (Matamala 2019: 66). Free commentary is a revoicing technique normally used during the live broadcasting of events and programmes that need to be adapted for a foreign audience. As opposed to interpreters, commenters do not have the necessity of translating the original voice track content. Rather, they can work by spontaneously adding and omitting information, or by offering insightful observations (Pérez-González 2014: 20). Free commentary is carried out in occasions such as international sport events, public debates, shows and ceremonies (e.g. the Oscars).

Finally, narration requires special attention. Indeed, a review of the academic literature reveals a lack of consensus regarding the definition of narration. Chaume (2012), for instance, defines it as "a kind of voice-over, where the translation has been summarized" (2012: 3). However, as Matamala (2019: 66) points out, rather than from a TS perspective, it seems more adequate to look at narration from a FS perspective. In this context, we will adopt the terminological distinction provided by Franco et al. (2010). The authors reserve the term narration to all the instances of "speech sequences by invisible speakers over programme images" (2010: 40), something purely related to the field of FS. Instead, they propose the label 'off-screen dubbing', to designate the activity of translating off-screen voices, entering into the domain of TS. For what concerns fiction,

this is the case of films in which the narration [4] of a third-person narrator[6] is translated into other languages [5], as in *The Killing* (Kubrick 1956), where a detached omniscient voice describes the event leading up to a robbery at a racetrack.

> [4] *Original English version*:
> "At exactly 3:45 on that Saturday afternoon in the last week of September, Marvin Unger was perhaps the only one among the hundred thousand people at the track who felt no thrill at the running of the fifth race".

> [5] *Italian version*:
> "Alle 15:45 di quel sabato dell'ultima settimana di settembre, Marvin Unger era forse l'unico tra le 100.000 persone presenti all'ippodromo a non provare alcuna emozione per le sorti della quinta corsa in programma".

On the other hand, in non-fiction genres, narration and voice-over often coexist. Documentaries, for instance, have sequences of spoken language from off-screen narrators that are faithfully translated (dubbed), and series of interviews in the SL which are voiced-over in the TL and still slightly audible in the background. (Franco et al. 2010: 42)

1.2.2.4. Audio description for accessibility purposes

Audio description (AD) is directly related to voice-over, as it also implies adding a supplementary voice track to an already multimodal ensemble. For this reason, it is also known as video-description. What distinguishes it from other revoicing methods is that AD falls into the category of assistive AVT methods, which means it is "intended to support viewers who do not have full access to some aspect of an audiovisual event" (Kruger 2020: 27). In particular, it is used to provide verbal descriptions of the visual component. This aspect makes AD an invaluable tool for the blind or the visually impaired persons (VIPs). These types of audiences, in effect, need to be assisted to fully comprehend and enjoy a film without its visual part (which contains a large number of signifying modes of a multimodal text)[7].

Unlike voice-over, the history of AD is rather recent, as the first audio description experiments for VIPs date back to the 1980s, when the technique was introduced in theatres in the US (Perego 2019: 116). AD is today especially used to enhance comprehension in film and television. Yet, since it can be applied to any multimodal and multimedial product or event that

[6] Analogously, this happens in films with first-person narrators – who may be the protagonists as in *Goodfellas* (Scorsese 1990), or minor characters – telling the story or expressing their subjectivity at various moments.

[7] Research on audio description focus on the best way to assist VIPs in the enjoyment of films. Kruger mentions an interesting strategy, that of 'audio introductions', where the audience is provided with additional contextual and stylistic information to optimize the overall experience that will follow.

incorporates a visual element, its scope is no longer limited to these domains. Indeed, it can also be employed in contexts such as sports events or museums, where artworks are explained to all types of visitors. Creating AD is, however, not an easy task, as describing visual elements to those who cannot see them is not the same as it is to explain them to those who can. Furthermore, AD has to cope with a number of technical restrictions, which will be discussed now.

Technically speaking, and for what concerns audiovisual content, AD verbalized visual information integrating it "with the existing auditory ensemble (consisting of dialogues, sounds, music, noise, silence, etc.) to form a new coherent text" (Perego 2019: 114). While other popular AVT methods such as subtitling and dubbing focus on dialogues, AD is not concerned in achieving direct equivalence to a source text (Kruger 2020: 27). Instead, a sort of intersemiotic translation occurs, as the focus is on transferring elements from the visual to the auditory domain. The rendering takes place between the stretches of spoken dialogues, and this time limitation leads audio describers to make serious choices about which elements are essential to convey and which can be omitted, in order not to overload the audience with excessive information (Pérez-González 2014: 26). The interaction between the visual and the auditory component needs to be well balanced, and even conveying basic (yet crucial) non-verbal information such as the look of the characters, their gestures or the setting of a scene can be particularly demanding[8]. In effect, as Perego (2019) summarizes, a good AD must be "'meticulous', 'concise', 'visually intense' and 'usable'" (2019: 119). On one hand, this means that the description must to be accurate and intense when describing important[9] visual details. On the other, audio describers should favour a plain and direct syntax, with a style that reflects that of the audiovisual product for the entirety of its duration. Another important aspect of AD guidelines concerns subjective descriptions, which should be avoided, if possible. When spectators watch films, their judgments are built upon their personal beliefs or inclinations. Similarly, objectivity must be pursued in AD to prevent "spoon-feeding or a patronizing attitude towards the target audience" (2019: 121), even though "the interpretation of gestures and facial expressions is sometimes essential in order to convey the filmic narrative as well as its aesthetic dimension" (Kruger 2020: 27).

[8] In addition to this, another problematic point is deciding whether to use technical cinema-related terminology to describe shots, angles, camera movements and other signifying directors' decisions, which, however, flow uninterruptedly throughout the entire film.

[9] When establishing the degree of importance of on-screen elements, professionals must take into account the audience's hypothetical background, i.e. "people born blind have no such visual memory to draw upon; hence, they have little or no interest in the colour of someone's hair, description of their clothing, etc." (Gambier 2018: 52).

1.3. Dubbing

Dubbing is an AVT technique that "consists of replacing the original track of a film's (or any audiovisual text) source language dialogues with another track on which translated dialogues have been recorded in the target language" (Chaume 2012: 1). As anticipated, however, the field is vast and requires special attention. This is true not only because dubbing is, together with interlingual subtitling, the dominant form of film translation, but also because, in the context of the present study, it is the AVT type chosen and analysed – in relation to new developments in the field of artificial intelligence. Before coming to the present, however, it could be beneficial to start by looking at the evolution of dubbing throughout its history.

1.3.1. Evolution of dubbing to the present day

Although used extensively for all products "destined for the screen that contain speech embedded in a multimodal context" (Pavesi 2020: 156) (including, for instance, video games) dubbing is inextricably tied to cinema and its history. The introduction of dubbing is a direct consequence of the commercialization of sound cinema, which began in the mid-1920s. Before, silent films could contain soundtracks and sound effects, but only if performed live by small orchestras (Cornu 2014: 24). The first feature-length movie with a synchronised sound system to appear in this context was *Don Juan* (Crosland 1926). Early sound-on-film technology, however, only allowed music and sound effects to be recorded. Just a year later, in 1927, engineers managed to record dialogues as well, and *The Jazz Singer* (Crosland 1927) was presented as the first ever talkie. With its 309 words of spoken dialogue, the film was a major hit and contributed to drastically redesign cinematographic landscape. "Image, word, music and background noise could now be used as representational resources to construct the diegetic world, hence opening up new avenues for artistic expressiveness" (Pérez-González 2014: 44). In addition to the artistic innovations that followed its introduction, sound technology soon pushed companies and distributors to find solutions to adapt films for foreign audiences, and strategies for translating spoken dialogues started to be elaborated for the first time.

Interestingly, sound films containing dialogue were the cause of a temporary financial crisis in the American film industry. Indeed, while the United States dominated the global film market in the 1920s, the advent of sound put a strain on American companies, which were unable to meet the demand for sound films in other languages (Pérez-González 2020: 31). Nevertheless, following the novelty of talking films, the abandonment of silent cinema was still quite slow. The slowness can be explained through two considerations. Firstly, silent films were still considered profitable and sustainable from an economic point of view. Indeed, translating and recording intertitles was easy and cheap. In the second place, there was a linguistic and cultural reason, because silent cinema

possessed a sort of universal language, a "visual Esperanto" (Chaume 2012: 11) that ensured the enjoyment of films to diverse audiences from various parts of the world. Actually, having to translate intertitles already meant that this potential was partly lost. Talkies' introduction, thus, constituted a new problem only in terms of technical work and costs.

Since earliest subtitled versions of American films proved unsuccessful in Europe, as many cinema-goers could not read yet[10] (Chaume 2012: 11), the first large-scale attempt to translate talking films consisted of the realisation of the so-called 'multilingual films'. In fact, multilingual films could be considered as a strategy to avoid translation, since they consisted of parallel versions of the story with different actors from other countries. These films could be regarded as ancestors of modern remakes, but they were not quite the same thing, as they contained sequences that could be used in all versions, and dialogue sequences that needed to be remade in other languages. The practice was, however, abandoned after a short period of time, because of the high production costs (2012: 2). Moreover, after the introduction of dubbing, multilingual films seemed to have lost their meaning, as audiences "were more inclined to watch dubbed North American films starring famous Hollywood names, rather than the multilingual versions that used the services of second-class actors and actresses" (2012: 12).

The birth of dubbing was quasi-incidental. In 1928, two engineers from Paramount Pictures managed to record a dialogue that matched the onscreen characters' lip movements (2012: 12). Initially, this process was elaborated to improve the quality of inaudible or poorly recorded dialogues, and was known under the name of 'post-synchronization' (Pérez-González 2020: 31) or 'doubling' (Pérez-González 2014: 46). It was soon clear that it would be possible to expand the use of this technique, for example by recording a translated version of the dialogues rather than another one in the same language. This constituted, in essence, the invention of dubbing as we know it, and is one of the most important milestones in the history of AVT.

The development and refinement of subtitling and dubbing techniques allowed American film industry to regain its central role by the mid 1930s (2014: 46), thus beginning the second wave of American domination over importing markets, which in that period included countries facing fascist regimes. Totalitarian governments of 20th century Europe viewed American domination with strong concern and growing hostility. On the one hand, this hegemony was a threat to national film industries. On the other, it constituted a menace to identity, language and culture, which had to be protected at all costs (Pérez-González 2020: 31). As mentioned earlier, Mussolini's nationalist policy prohibited all instances of foreign language in cinemas in the 1930s. In this context, dubbing was embraced as the perfect weapon to protect the country from the alleged overseas evil ideologies

[10] In addition to economic factors and the size of the film industries, literacy levels strongly contributed in shaping the map of countries that prefer subtitling over dubbing. This is why, in the Scandinavian countries, where most people could read, subtitling was adopted without resistance.

that could penetrate the audience's way of thinking. The 20th century is the perfect demonstration of how powerful mass entertainment and translation can be (or be perceived) in controlling the lives of citizens. In effect, fascist Italy paved the way for many other oppressive regimes, like Francoist Spain and Hitler's Germany, which "greeted this measure with enthusiasm and adopted it in their own countries some years later (Germany, the Reich Film Law, 1934, and the Enabling Act, 1936; Spain, Act of 23 April 1941)" (Chaume 2012: 13). As far as production and distribution dynamics are concerned, the efforts and the heavy linguistic prohibitionism did not changed the situation, as American domination did not cease (Pérez-González 2020: 31). For what concerns the favourite AVT method in former dictatorship countries, however, politics and ideology played a major role, contributing to the creation of deep-rooted habits difficult to eradicate among the population. "In Spain, for instance, […] this modality of AVT remains strong nowadays insofar as Spanish viewers have become used to dubbing and what it involves" (Bosseaux 2019: 49). Similarly, dubbing became a specialty in Italy, that then took root well beyond fascist policies (Pavesi 2005: 20).

Academic texts on AVT traditionally establish clear-cut distinctions between countries and their associated AVT method. For what concerns Europe, referring to the division proposed by Chaume (2012), one could divide the map between dubbing (e.g. the so-called FIGS countries[11]) and subtitling countries (e.g. Albania, Croatia, Denmark, Finland, Greece, the Netherlands, Norway, Portugal, UK). A third category includes countries that use both AVT modes, such as Belgium, where traditionally dubbing is used in Wallonia and subtitling is used in Flanders (2012: 6)[12]. This was already partly mentioned in Section 1.2.1. What was not said, however, is that this simplistic distinction is no longer entirely valid as of today. Chaume himself, after having outlined the map, states that the "AVT landscape is no longer black and white" and that "the distinction between dubbing and subtitling countries have become blurred" (2012: 7). Chaume also provides several reasons why the scenario has changed, and facts that refute clichés about supposed clear boundaries. To give another example, subtitled films are now shown on a daily basis in many cinemas in traditionally dubbing countries for new types of audiences (2012: 6-7). Similarly, drawing upon other studies, Matamala et al. (2017) talk about the new global AVT landscape adopting a 'sociological approach', which emphasizes the importance of many factors in the choice of the transfer mode, including the film type, the intended audience, and the environment where the film (or the alternative audiovisual content) will be projected. For example, "popular films for wide audiences will be dubbed and released in big cinemas; popular films with artistic quality will be

[11] France, Italy, Germany and Spain.

[12] Voice-over countries also exist (e.g. Poland, Russia, Bulgaria, Latvia, Lithuania), although times are changing for them, as they are shifting to more widely adopted AVT methods as well.

dubbed and subtitled, and released both in big cinemas and art houses, and "auteur" films will only be subtitled and released in festivals or art houses" (2017: 4).

Since text type is a crucial variable to be taken into account, it seems now interesting to mention cartoons[13], to understand the role of dubbing in today's media landscape. The reason behind this relationship is that "cartoons for younger children are dubbed all over the world" (Chaume 2012: 2), even in traditionally subtitling countries. Thus, this audiovisual text type, in the first place, refutes the standardised distinction between subtitling and dubbing countries. As it was observed at various times so far, the reasons are both of technical and economic nature. Technically speaking, dubbing is generally preferred for children's cartoons because the targeted audience might have insufficient reading skills and might face problems in the comprehension (O'Connell 2003: 223). Another technical reason is that typical problems of lip synchronisation are limited in animated films compared to films with human actors (2003: 223). The topic of synchronisation is crucial for dubbing and will be discussed later in relation to technical constraints associated with the technique. The second reason why cartoons are dubbed worldwide is of commercial nature. Indeed, "the fact that high quality animation can be revoiced for rebroadcast to a new audience of children at a fraction of the total original production costs […] makes dubbing an attractive option" (2003: 223). Moreover, although mainly aimed at children, animated films "are conceived in such a way as to appeal to adults as well and they have different layers of meaning" (Minutella 2021: 2). Given their wide and diverse audience, animation is in vogue today, and there currently is an exponential growth in the production of such text types (Sánchez-Mompéan 2015: 270). To attract more viewers (and generate more revenue), production and distribution houses take advantage of dubbing to devise strategies that are exclusive to animated films. "Apparently, one of the most effective ways of drawing a wider audience consists of casting famous personalities to lend their voice to animated characters" (Sánchez-Mompéan 2015: 271). As testified by the memorable performance of Antonio Banderas in the interpretation of Puss in Boots in *Shrek 2* (Adamson et al. 2004) or Jack Black's iconic characterisation of Po in *Kung Fu Panda* (Osborne and Stevenson

[13] In fact, the following considerations apply to cartoons (understood as animated products for children) but also to animated films in general. Indeed, 'animation' is an umbrella term encompassing a number of techniques, such as traditional animation (the oldest and perhaps most time-consuming, because frames are hand-drawn one by one as in old Disney's films or most Japanese Studio Ghibli's anime), 2D digital animation (where computer software are used to generate intermediate frames and fluid animations in a two-dimensional space) and 3D animation, more complex because it adds a third dimension, with Pixar's *Toy Story* (Lasseter 1995) as the best example, being the first feature-length film made entirely in CGI (computer-generated imagery). Other animation techniques can be very different. Motion capture, for instance, is a technique where real actors wear bodysuits and face-masks with special sensors which closely capture their movements. Data is then processed by animators who use them in the production of video games or films such as Zemeckis' *The Polar Express* (2004), the first film to use this technology. Stop motion is another versatile animation style used in box-office hits such as *Chicken Run* (Lord and Park 2000). The technique consists in meticulously capturing series of still images where physical objects or puppets are brought to life through the illusion of movement. The list is not exhaustive as other styles also exist and many films make use of mixed techniques. To avoid ambiguity, we will employ the term 'cartoon' to designate specific children-targeted products, whereas the more general term 'animated film' will be used to encompass the whole range of different techniques mentioned above.

2008), this is a particularly relevant phenomenon in the US. The same happens with dubbed versions of American films in other countries, or with domestic productions, as in the case of the Italian film *La gabianella e il gatto* (D'Alò 1998), which included the voices of Carlo Verdone and Antonio Albanese, who are famous actors yet not professional dubbers. This commercial strategy not only tells us why animated films are extensively dubbed, but is also an important proof of the vitality and versatility of dubbing as a transfer mode in the international market.

In Section 1.2.1., we discussed the expansion of subtitling in relation to new contexts of consumption of audiovisual contents. While this trend is now well-documented, the parallel spread of dubbing as a practice is perhaps under less spotlight. In particular, another notable phenomenon testifying dubbing's increase use is fandubbing. As Chaume (2012) puts it, if fansubbing is "the domestic subtitling by fans of TV series, films or cartoons (especially anime) before they are released in the fan's country", fandubbing is its lesser known yet exact counterpart, i.e. the "domestic dubbing of trailers and cartoons that have not yet reached the fans' country" (2012: 4). Even if the practice raises issues in terms of copyright violation, the phenomenon is quickly spreading across the world, investing even traditionally subtitling countries (Chaume 2013: 117). Thanks to the immediacy provided by new user-friendly technological tools[14], today users can freely experiment and broaden the traditional uses of fandubbing, which is sometimes alternatively spelt 'fundubbing' when referred to the practice of revoicing audiovisual content for humoristic purposes (e.g. for the creative production of viral memes on social media) (Chaume 2012: 4).

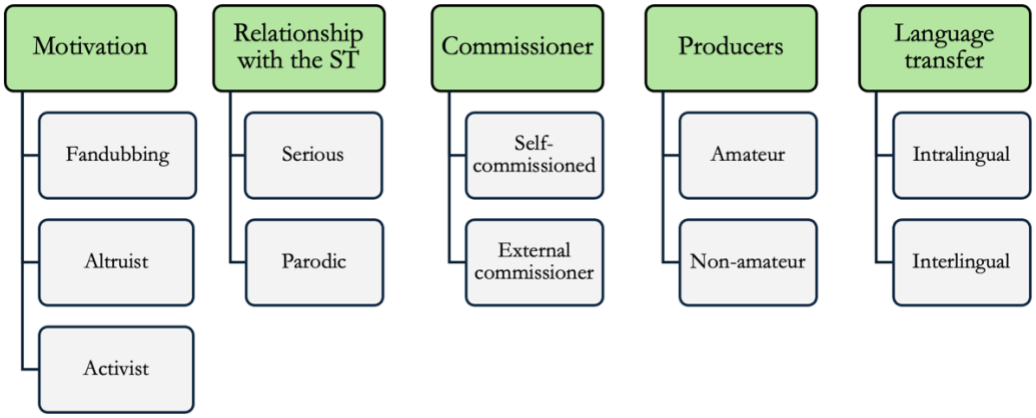
Amateur (humoristic) dubbing is only one realisation within the macro-area that Baños and Díaz-Cintas (2023) defines as 'cyberdubbing'. The authors propose this label as an umbrella term to address to "the complexity of the multitude of practices found on the internet" (2023: 134) and to encompass a wide range of dubbing types which "are no longer performed by 'fans' and […] differ significantly from how these practices originated back in the 1980s" (2023: 130). *Table 1* visually summarises the classification[15] proposed by Baños and Díaz-Cintas, which is divided into various levels. The first level consists of the identification of the dubbing motivation: apart from the already known 'fandubbing', the authors mention 'altruist dubbing' (i.e. made, for instance, for educational purposes), and 'activist dubbing' (e.g. when dubbing becomes a weapon for satire or political criticism). The second level focuses on the type of relationship between the dub and the source text, which can be either 'serious' or 'parodic'. Two further levels include the distinction between those who produce and those who commission the actual dubbings: on one side, producers can be 'amateurs' or 'non-amateurs' (trainees or professionals); on the other, the work

[14] Many of these dubbing tools can be accessed online for free. Therefore, in theory, all is needed for the creation of most amateur dubbings is a device with a microphone (either a computer or a smartphone), an internet connection, and one's own voice and creativity.

[15] Baños and Díaz-Cintas reiterate this classification in its entirety also with regard to subtitling, considering "the myriad new subtitling types that have emerged in the mediascape in the past decades" (2023: 131).

can be 'self-commissioned' or commissioned by others. Finally, it is necessary to consider the relation to language transfer, given that cyberdubs can be either "intralingual" or "interlingual".

Table 1: cyberdubbing classification by Baños and Díaz-Cintas



To conclude, it is important to mention the new phenomenon of automatic dubbing. In effect, similar to what happened for subtitling with respeaking, automation has entered in recent years into the field of dubbing as well. Automatic dubbing consists of the generation of synthesised translated speech, without the need of human dubbers, and using AI-based technologies. Since this constitutes a fundamental topic of the present work, a more detailed explanation will be provided later on.

1.3.2. Dubbing is not dead: the rebirth of English-language dubbing.

There is scarce academic attention on the phenomenon of English-language dubbing. This is certainly due to the fact that "English-dubbed films are a rarity in English-speaking countries" (O'Sullivan and Cornu 2019: 23). The film industry has been dominated by the US for a long time now, and the flow of audiovisual products has mainly been from anglophone countries to the rest of the world. Therefore, it is legit to state that "film translation into English […] is, after all, translation against the current of the film trade" (2019: 23). This phenomenon will be, however, one of the main subjects of analysis in our work, since it constitutes an entirely new scenario which is very relevant to today's AVT state of the art.

There have been various historical moments in the past where foreign films' have achieved notable success in English-language markets[16]. Particularly, "key moments […] include the early years of sound […] and the period immediately following the Second World War" (O'Sullivan and Cornu 2019: 23), when American market witnessed the import boom of Asian and especially

[16] Of course, foreign films are to be considered as such only if looked from an anglophone perspective.

European films. New movements (both artistic and commercial), such as the *Commedia all'italiana* or the French *Nouvelle Vague*, appeal to a new, much-changed audience of young people and generally well-educated cinema-goers. Another popular trend from Europe was the subgenre of Spaghetti Western (i.e. westerns produced and/or directed by Italians), of which Sergio Leone was the greatest exponent. As Pérez-González (2014) explains:

> "Leone's films belong to that rare breed of films shot in a foreign language and later dubbed into English for the enjoyment of mainstream American audiences. In his westerns, all dialogue was routinely re-recorded during the post-production stage for a number of reasons. For instance, the limited affordances of the cinematographic technology available in the 1960s […] prevented the Italian director from using synchronized sound to shoot a large number of scenes […]. On a different note, Leone's films often involved multilingual casts, with actors performing and addressing each other in their respective languages" (2014: 32).

Leone's films were much criticised for their use of sound and dialogue, because they challenged "Hollywood's steadfast adherence to synchronous diegetic sound – which […] became a means to construct characters with an unambiguous voice, and hence to articulate a national identity based on individualism and meritocracy" (2019: 33). The 'Old Hollywood' of the early 1960s, on its part, struggled to evolve. In effect, studios faced economic difficulties, since traditional productions (e.g. musicals or historical epics which required huge budgets) did not meet the new audience's expectations and desires. From the 1970s, and with the explosion of the 'New Hollywood' (a new generation of directors capable of experimenting with less conventional choices in technique, narrative and subject matter), North American cinema regained its position and foreign films returned to have much less impact on the box office (O'Sullivan and Cornu 2019: 23).

Today, the situation has completely changed. In the last section, fandubbing and animated films have been mentioned as evidence of dubbing's viability. However, the main reason behind dubbing's take-over of significant parts of the audience in traditionally subtitling countries is to be found elsewhere. Indeed, it is reasonable to affirm that "the increasing digitization of audiovisual commodities witnessed since the middle of the first decade of the twenty-first century led to the fragmentation of national audiences and the formation of geographically dispersed audienceships" (Pérez-González 2020: 31). This has boosted the practice of both general interlingual dubbing and English-language dubbing, which was a rather marginal phenomenon until now. Thus, on one side, English-language films are distributed quickly all around the world; on the other, the demand (and

the production) of local[17] films and TV series is higher than ever. In this context, mobile devices and streaming platforms enable audiovisual texts to be enjoyed from an individual perspective in a taylor-made experience based on personal preferences.

The streaming giant Netflix offers an exemplary opportunity to analyse the phenomenon. As a matter of fact, after years of increasing success, the Californian company has been recently confronted with rising competition and slowing subscriber growth (Lee 2022). Surprisingly, to diversify its own business model, Netflix identified viewers' increasing interest in local productions and "made the decision to start producing locally to export globally" (Bonella 2023: 4). By releasing dubbed and subtitled version of original local productions, in effect, platforms seek to appeal to wider segments of the audience and to maximise the economic return for their previous investments (Pérez-González 2014: 158). The company directors have declared that, only in 2021, viewers of dubbed content increased by 120% (Lee 2022), and that figures show how dubbed versions of popular local productions are frequently more watched than their subtitled counterparts, so much so that someone started talking about a "dubbing revolution" (Roxborough 2019). One reason behind this explosion has been the surge of South Korean audiovisual products that have captivated world audiences. The most emblematic case is undoubtedly the dystopian thriller *Squid Game*[18] (Hwang 2021-), which is to date the platform's biggest TV show ever (Bonella 2023: 5). Other examples also come from Italy. Zerocalcare's show *Strappare lungo i bordi* (2021), for instance, has crossed national borders and has received attention online for being poorly dubbed in other languages (e.g. Gastaldi 2021). In fact, in recent years, this Netflix-driven revival has been accompanied by a great deal of discussion about the quality of English-language dubbing. In online spaces such as Reddit, threads have multiplied in which questions [6] are raised by users who are not familiar with dubbing.

> [6] "English Dubs […] are appallingly bad on Netflix. Is it me? Did I not realize until now just because of sheer volume of media content? […] Why are all the English voice actors so… boring?" (u/MorganAndMerlin 2023).

As Sánchez-Mompeán (2021) clarifies, if sometimes English dubs' quality is actually poor, on the other side much of this criticism can be explained by the lack of habituation to dubbing by the English-speaking viewers, who "perceive the dubbed dialogue as weird, emotionless or 'too dubby' if they compare it to the non-dubbed fictional speech they are used to consuming" (2021:

[17] The term "local" is used here interchangeably with "foreign", to refer to original productions in countries other than the United States.

[18] Original title: *오징어 게임* ; Romanization: *Ojing-eo Geim.*

185). On its part, Netflix is betting high in English-language dubbing, and this is testified by the redubbing[19] of famous shows such as the Spanish success *La casa de papel* (Pina 2017) (2021: 188).

English-language dubbings are today also created by users on social networks in the context of fandubbing. For this reason, the phenomenon constitutes a central theme for the present work, as extracts from Italian films dubbed into English will be taken into account to conduct the reception study.

1.3.3. Technical process and difficulties

As already outlined above, the term dubbing denotes the replacing of the source voice track with dialogue recorded in a different TL. From a technical viewpoint, however, this is a long and complex procedure that includes a number of different skills. We will now refer to the summary provided by Chaume (2012: 29-36) to describe the various stages of the creative process and the work that professionals do. A dubbed film is, indeed, the sum of many specialists' work that follow a specific production chain. The process begins with translators who adapt the original script to the target language in question. The translated text is then synchronised with the on-screen images (which include the lip movements of the actors, but also many other aspects that constitute the main problems associated with dubbing, as we will see later) by dialogue writers. In addition to this task, these figures take also care of adding other information (e.g. paralinguistic features) to the text, in order to help dubbers during their performance.

There is terminological confusion among the terms designating the professionals who dub films. Indeed, with regard to interlingual subtitling, the terms 'voice actor', 'voice artist', 'voice talent', 'dubbing actor', 'dubbing artist' and 'dubber' are used roughly interchangeably. However, a distinction between some of these labels becomes necessary when considering the genre of animation, where both inter- and intralingual subtitling are to be found. Quite often, outside the academic context, "the task of giving a voice to animated characters in the language spoken by the country where the feature film has been made" (Sánchez-Mompéan 2015: 274) is defined as voice-over. Nevertheless, as already stated in Section 1.2.2.2., there are differences existing between dubbing and voice-over that preclude this comparison. On the other hand, as suggested by Sánchez-Mompéan (2015), when there is a translation dimension involved (interlingual dubbing), it is possible to use the term 'dubber' or another one from the list above. On the contrary, it would be appropriate to speak of a 'voice actor' only to indicate an actor who lends his voice to a character in the process of intralingual dubbing, which does not involve any language switching.

[19] Redubbing is a complex and expensive technique that consists in the creation of new translations and new recordings with different voice casts, aiming at raising overall quality of the dubbed version.

Dubbers are carefully selected by dubbing directors, who assist them throughout the entire creative process, just as film directors do with actors on stage. The whole production chain is carried out in studios and supervised by skilled technicians and sound engineers. The large number of agents involved in the dubbing process only adds to its complexity, especially considering that technique-related constraints are also accompanied by more general issues concerning the translation process *tout court*, which is already challenging by itself.

Linguistic adaptation for dubbing is, indeed, a task of utmost difficulty, since the translated dialogue needs to take into account several para- and extralinguistic elements, including actors' performance and movements, settings, sounds and background noise, to name a few. Considering this, while sometimes referred to as 'total translation', dubbing is defined by Pavesi (2005: 12) as a type of 'constrained translation', as elements which are embedded in the film constitute additional difficulties for translators. Most of academic research on dubbing has been done since the late 1980s and has focused on the analysis of its medial constraints (Bosseaux 2019: 50).

Synchronisation, i.e. the correspondence between a visual and an acoustic element, is universally acknowledged as being the main one. Pavesi (2005: 13-16) describes various types of synchronisation, mentioning the categorisation provided by Herbst (1994), who spoke of lip synchronisation and nucleus synchronisation. The author changes the label of the first category from lip to articulatory synchronisation, because it is essential for translators to consider not only the lips but also the movements of other articulators involved in the phonation process, such as the jaw[20]. Chaume (2012: 74) explains that, in some cases (e.g. when film scenes include close-ups) this audio-image equivalence has the priority even over semantic and pragmatic equivalence. Yet, articulatory synchronisation is not crucial in all cases. As a matter of fact, text-type has an impact on the degree of importance of synchronisation (Bosseaux 2019: 51). As already mentioned before, animated films can have flexible lip synchronisation. Sometimes this is related to the intended target audience, because, in the case of children, they are not particularly demanding. Moreover, in a broader sense, this can be related to animated characters' simplified physical representation that reduce the need of accurate synchronisation (Chaume 2004: 45). Take for example Wes Anderson's stop-motion films (i.e. *Fantastic Mr. Fox* 2009, *Isle of Dogs* 2018). In such cases, it is possible to see how puppets exhibit some sort of articulatory movements that roughly match the dialogue lines. Nevertheless, the amount of work required in revoicing the almost random lip movements of animated characters is not comparable to the accuracy needed when human actors are involved. In contrast, nucleus synchronisation is essential in the aforementioned textual genres as well as in film translation in general. Pavesi (2005: 25) describes it as paralinguistic and kinetic synchronisation,

[20] Articulatory synchronisation is further divided into quantitative (i.e. the correspondence of the dubbed speech with the beginning and end of the original actors' speech) and qualitative synchronisation (i.e. the actual matching of the sounds of the dubbed speech with the visible articulatory movements).

i.e. the matching between speech and body movements[21] which all serve to confer emphasis and express communicative salience.

Another problematic aspect to be taken into consideration is the treatment of multilingualism in films. Multiple languages can be present within the same film for multiple reasons: they can be part of the linguistic landscape (used, thus, as indicators of a multicultural setting), or spoken by individual actors (for characterization and/or for plot necessities). These choices can unlock great narrative potential for filmmakers, but at the same time, « continue to pose a challenge in the context of dubbing » (Pavesi 2020: 160), as the many studies about this phenomenon testify (e.g. Heiss 2004; Abril 2015). In some cases, no linguistic hints pointing at the diversity of the languages spoken are present in the dubbed version[22]; in other cases, foreign accents can be used in the target language; alternatively, the main source language can be translated into the desired target language, while all the remaining languages are kept without modification. Baldo (2009) signals another recurrent strategy used in the translation of films containing Italian as 'exotic' language. In the Italian versions of such films, while the main source language is replaced by standard Italian, source-text Italian is instead replaced by a dialect. A well-known example is the much-discussed translation of a scene from Tarantino's *Inglorious Basterds* (2009), in which a group of undercover Jewsish-American soldiers confront a Nazi officer, while pretending to be Italian. In the Italian version of the film, the translation process resulted in the group speaking Sicilian dialect in front of the officer (who, in turn, conversed with them in Italian).

Besides these issues, the audio-visual relationship typical of filmic translation continuously raises problems of different orders. For instance, what in the source culture is only visually shown, often requires verbal codification in the target culture. It is the case of cultural, humorous or symbolic elements which are embedded in the visual dimension of the work, which are immediate or implicit for an audience but requires explicitation for another. Hence, filmic translation for dubbing is, of course, a type of interlinguistic transfer, but also frequently an example of intersemiotic translation (Pavesi 2005: 16).

1.3.4. Features of dubbed language

Pavesi (2005: 28) mentions the term 'dubbese'[23] to designate the language of dubbing as product of the filmic translation. The term is often used with a derogative connotation, since many studies have highlighted the artificial nature of the language of dubbing, which can differ significantly from

[21] Hand gestures, facial expressions, head movements, eyebrow and eye movements, and so on.
[22] This can create paradoxes when different languages are named in the dialogues but not spoken by the characters.
[23] In fact, '*doppiaggese*' (in the original Italian version) is used to refer to the peculiarities of translated filmic Italian. Nevertheless, the considerations that follow are believed to be true for the language of dubbing in general.

the language of the filmic dialogue[24] in the source text, which, in itself, already has different premises compared to spontaneous speech. Indeed, language use is carefully measured and planned in films. This is true because dialogues contribute to shape the artistic value of the audiovisual product, and they entertain the audience. However, more importantly, dialogues are means used to condense information and narrate the story, portraying characters and situations. For this reason, they are semantically dense, typically explicit and punctual; on the contrary, spontaneous conversations are less structured and contain implicit content (which belongs to interlocutors' previous knowledge) (2005: 31). Nevertheless, it would not be wrong to consider filmic dialogue in the source language as a faithful representation of spontaneous speech. First, a certain evolution in the language of filmic dialogue did occur. Whereas in the early stages of film history dialogues were similar to theatrical exchanges due to their evident artificiality, today they show traits which are characteristic of spontaneous speech (2005: 32). Second, despite the assumptions made above exist, films present a multitude of different discursive situations in which there are attempts to portray authentic orality (2005: 30). For this reason, corpora containing film dialogues (e.g. *Pavia Corpus of Film Dialogue*) are often used in research for conversation analysis purposes.

Let us now turn to the language of dubbing, which requires different considerations. It might be useful to analyse the typical features of dubbese through three levels of linguistic analysis, i.e. phonology, morphosyntax and lexicology.

As far as phonology is concerned, one of the main universally recognised traits is the absence of diatopic variation[25], which seems to be an inevitable consequence of the linguistic transfer. Indeed, many have highlighted the difficulty (or the impossibility) in finding varieties having equivalent cultural and symbolic values as the geolects spoken in the original version of a film (Pavesi 2005: 36). In this sense, many examples can be found in North American or British films dubbed into Italian. In Kubrick's *Dr. Strangelove*[26] (1964), an unhinged general of the US Air Force confronts a British officer of the Royal Air Force. The Italian version, despite an excellent dubbing that succeeds in expressing the comic yet meaningful exchange, has no way of replicating the original version's accent differentiation. Similarly, Guy Ritchie's *The Gentlemen* (2019) features a cast of actors who make use of different English accents to characterise a group of gangsters with a varied background. In the original version, it is possible to distinguish Matthew McConaughey's American accent, Hugh Grant's London working-class Cockney accent and Colin Farrell's Dublin

[24] With "language of filmic dialogue' we refer to the linguistic realisations in a source text. The 'language of dubbing', on the other hand, is the product of interlinguistic translation and revoicing of such text.

[25] It is well known that language can vary according to regional, social, or contextual differences. In sociolinguistics, a specific variety may also be called a "lect". Regional dialects (i.e. varieties of a language spoken in a given geographical area of a country, sometimes significantly different from standard varieties and considerable as languages on their own), sociolects (associated with particular classes and groups of people), or idiolects (i.e. individual varieties, because every person has a unique and different way of using language), are all examples of lects, that can vary in terms of phonology, morphology, syntax, lexicon, or other levels of analysis.

[26] Full title: *Dr. Strangelove or: How I Learned to Stop Wowrying and Love the Bomb*.

accent. However, in the Italian version, the geographical provenance of the characters is impossible to discern just by hearing them speak. In the new media landscape, a similar 'culture-neutralising' force can be found as a consequence of recurrent production and marketing strategies. Indeed, "producers from 'marginal' [non-US] film industries around the world have tried to transcend the barriers of the art house circuit by imitating US film production values and, increasingly, by borrowing English as their shooting language" (González Ruiz and Cruz García 2021: 222). In Europe, this seems to be a much-consolidated trend today, with directors choosing to shoot in English and using international casts. Well-known non-US productions worth mentioning are, among the others, the recent *The Room Next Door* by Spanish director Almodóvar (2024), *A Bigger Splash* by Italian director Guadagnino (2015), or Swedish director Östlund's *Triangle of Sadness* (2022). Since this work is concerned with English-language dubbing, it may be useful to underline that in these and many other international[27] productions, English is the main source language, meaning that translated versions are rarely constrained by the difficulties of rendering language varieties specific of the country of origin of the film. On the other hand, it is specifically when marginal[28] films display high degrees of language variation that dubbing is perceived as an element that "prevents the audience, at least partially, from becoming aware of the distinct idiosyncrasies of the people and the places depicted in the screenplay" (González Ruiz and Cruz García's 2021: 219). Original Italian films, especially, traditionally play with the potential of the rich national sociolinguistic landscape[29] for characterisation purposes. Despite historical marginalisation[30], for instance, dialects have always been used in Italian cinema. This seems especially true if we consider, for example, the use of southern dialects in Italian comedy films[31] (Romano 2015). While there are

[27] Here, 'international' is not referred to the production stage, but rather to the distribution goal. The films are, thus, non-US domestic productions which nevertheless aim at overcoming the national barriers inevitably imposed by filming in the national language (and that would, at least, require the film to be subtitled into English).

[28] By 'marginal' we mean non-English language national productions.

[29] Berruto's (2012) scheme offers a reliable overview of the architecture of contemporary Italian. On a multidimensional continuum, Berruto traces three axes of variation, which represent three different dimensions in which a language can vary: the diastratic dimension, the diaphasic dimension and the diamesic dimension. The first dimension concerns the variation in relation to the social background (e.g. age, sex, education) of a speaker; the second dimension is related to the situation and the setting in which the communicative act takes place; the third dimension refers to the variation on the basis on the medium adopted for the communicative act (e.g. written vs. oral). While present in the first 1987 architecture, diatopic variation does not appear in the updated version of 2012. The author explain that this does not mean that it is less relevant, but rather, that it is not graphically shown as it is so important for the Italian sociolinguistic landscape, that it is always present in the background (and it always has to be taken into consideration).

[30] Negative attitudes towards dialects originated during Fascism (Raffaelli 2010) and continued for a long time. However, some scholars have revealed that evaluations are changing in recent decades. Italians do use dialects alternating it with Italian in the same communicative act (Berruto 2007: 145) according to code-mixing and code-switching modes (Antonelli 2016: 32). The renewed perception (and widespread use) of dialect is also evidenced by its incursion into previously unexpected domains, including advertising (e.g. Matrisciano 2021), politics (e.g. Molinari 2024), music (mostly through rap music, an intrinsically experimental genre from a linguistic point of view), and new media (e.g. Palermo 2023), where its presence seems to have increased exponentially.

[31] Indeed, examples of iconic southern comic characters can be abundantly found in the past: take, for instance, the memorable Neapolitan duo Totò and Peppino in films such as *Totò, Peppino e… la malafemmina* (Mastrocinque 1956). Still, this seems like a well-established custom of Italian cinema, as southern dialects are used as comic

ways (exemplified below) of rendering sociolects, diatopic differentiation is more complex to achieve, because "the connotations of different target culture dialects will never be the same as those of the source culture dialects they replace" (Díaz Cintas and Remael 2007: 191). To confirm this generalisation, it is possible to mention Bonella's (2023) study on some Italian Netflix original series dubbed in English. The research highlighted that strategies for rendering Italian dialects were almost entirely related to morphology, syntax and word choice, while no attempt of replacing the original dialect with another existing English dialect or variety was made. This last strategy is often viewed as a "not-so-politically-correct decision" (Bonella 2023: 6), since it might lead to confer (perhaps negative) social and cultural attributes to other varieties on the basis of individual inclinations or common sense. Nevertheless, this treatment of diatopic variation is still found in contexts that constitute some exceptions. Instances of this approach are to be found in animated films or TV series, which use it to obtain humoristic effects. Dialects and (internal or foreign) accents are often used in the animation genre to "trigger connotations and stereotypes in the audience's mind" (Minutella 2021). A famous example is the Italian version of *The Simpsons* (Brooks et al. 1989-), where many of the original version's accents are replaced by a constellation of Italian regional dialects[32].

Diastratic variation, on the other hand, is perhaps more easily achievable through a careful use of morphology and syntax. Pavesi (2005: 38-42) exemplifies various morphosyntactic strategies useful to recreate the effects of certain sociolects on the audience, i.e. right and left dislocations, cleft sentences, and non-standard use of verb tenses. These and other compensation strategies can be found, for instance, in dubbed versions of films containing varieties such as AAE (African American English), as in the excerpt below from Spike Lee's iconic *Do The Right Thing* (1989). The character Buggin' Out (marked with B) has an incident with a passer-by in his neighbourhood, and together with his friends like Ahmad (marked with A), he argues with the stranger about the matter [7]. In the Italian version [8], the dubbed lines contain examples of pronominal particle's elision[33] and right dislocation[34], morphosyntactic choices aimed at reproducing a lively and colloquial speech register.

devices in contemporary successful comedies too, as in the case of *Quo Vado?* (Nunziante 2016), featuring Apulian comedian Checco Zalone. *Quo Vado?* is currently the highest-grossing Italian film in Italy. Zalone co-wrote and starred three other films which are, respectively, the second-, third- and fourth-highest grossing Italian films in Italian cinema history (see https://movieplayer.it/film/boxoffice/italia/di-sempre/). This certainly confirms audience appreciation of Zalone's films (which are based on a clever stereotypical characterisation of a southern character, and where the regional variety spoken plays a major role), but also the inextricable link between Italian cinema and the highly characterising power of diatopic varieties.

[32] These include, to name but a few, Chief Wiggum's Neapolitan accent, Groundskeeper Willie's Sardinian accent, Carl's Venetian accent, and Reverend Lovejoy's accent, which is situated between Sicilian and Calabrian.

[33] "*M'hai*" instead of "*mi hai*".

[34] The adverb "*proprio*" can be considered as a right dislocation, since it occurs outside the clause boundaries on its right.

[7] *Original English version*:

B: "Not only did you knock me down, you stepped on my brand new white Air Jordans that I just bought and that's all you can say, 'excuse me'?"

…

A: "Yo, man, your Jordans are fucked up!"

[8] *Italian version:*

B: "Non solo m'hai distrutto una spalla, m'hai anche pestato le mie Jordan nuove di zecca appena comprate, e non sai dire nient'altro che scusami?"

…

A: "Ehi, fratello, le tue Jordan sono fottute, proprio!"

Given the difficulties in reproducing the phonological traits of a source language in a dubbed target language, it is crucial to adopt alternative strategies in order not to completely neutralise varieties that have been chosen for specific reasons in the original work. Each variety, in effect, carries a certain cultural load, prestige, stereotypes and expectations. Initially, Whites considered AAE as inferior, because it deviated from the grammatically correct standard. In fact, it is a variety that exhibits distinctive phonological, morphological, syntactic, semantic and lexical patterns, which is why it is internally structured and coherent. Many have pointed out how the use of this variety actually constitutes a bold example of the reappropriation of power, voice and identity (e.g. Taronna 2016). Therefore, the translator/adapter must be conscious of two key aspects when attempting to reproduce the source text's linguistic varieties. First, neutralising them could result in the loss of several nuances in the characterisation of characters and in the overall ambience of the film. Second, it may cause the failure in conveying the linguistic dynamics (of power or discrimination) of the universe depicted in the film to the foreign audience. Indeed, it has been shown that language is often employed as a means of maintaining or establishing inequalities within a society[35] (e.g. Lippi-Green 2012). Hence, maximum attention is required in the whole translation process, as often "the total effect of a series of microstructural shifts may just as well amount to a significant shift on the macrostructural level" (Delabastita 1990: 103).

Lexicon is the most superficial and perhaps most evident layer of a language. For this reason, its importance should not be underestimated in the representation of language varieties. As pointed out by Pavesi (2005: 42), lexicon is commonly perceived as a weak spot in the language of dubbing, since word choice often leads to unnatural or bizarre expressions which are at least marginal in common usage. Typical features of dubbed language concerning vocabulary are the toning down

[35] Lippi-Green talks about 'linguicism' to designate the linguistically argued racism which occurs particularly against those who speak variations of the standard language. Linguicism may be covert (unconscious and passive) or overt (conscious and active), and even systemic (when discriminations become rooted in the sociolinguistic scenario of a country).

of obscene language (Formentelli and Ghia 2021), the neutralization of terms that have strong stylistic or cultural connotations in the original, or the adoption of semantic or structural calques[36] from English. This often results in the emergence of so-called translational routines, source-language oriented translation solutions that are now common in most dubbed films (e.g. the typical Italian dubbese expressions 'Ci puoi scommettere' or 'Dacci un taglio', from the English 'You can bet' and 'Cut it out' respectively) (Pavesi 2005: 48-50).

In 1995 Lawrence Venuti published *The Translator's Invisibility. A history of translation*, an important (and much debated) text for Translation Studies, in which he compared two opposing approaches in the practice of literary translation, i.e. domestication and foreignization. The terms refer to the degree to which translators make target texts conform to the target culture. Domestication, on the one hand, may involve the loss of certain information and details, but it also puts the audience at ease, facilitating their ability to decode the work. What is often referred to as "neutralization" therefore falls under the realm of domestication, which ultimately aims at fluency as the most important quality for a translation. Venuti, on the other hand, thinks that the removal of all traces of foreignness in order to achieve fluency and transparency result in an actual "ethnocentric violence of translation" (1995: 20), and advocates for the foreignization strategy (which consists in experimenting and breaking conventions in the target language to construct a sense of otherness from one's own cultural norms, canons and ideologies). In the case of audiovisual translation, and for what concerns the language of dubbing in particular, the features described above show that target texts are generally 'domesticated', causing a distorted perception of the works on the part of the audience (even if this view sees the audience as a completely passive participant in the fruition). In general, Pavesi (2005: 59) underlines how dubbese seems to match the laws proposed by Toury in 1995, i.e. the law of growing standardization and the law of interference. Indeed, similarly to the domestication strategy, the law of growing standardization states that in translation, textual relations and features of the source text are replaced by less creative and more habitual (hence, standard) options in the target language (1995: 268). The second law stipulates that interference from a source language (which applies to the previously mentioned translational routines but also to other phenomena) is a default process in translation.

In light of all these issues, it would seem almost difficult to enjoy dubbed films. However, while it is true that translation (and interlingual dubbing included) always entails the possibility of losing some important aspects that are difficult or even impossible to transpose into a target

[36] A semantic calque is an expansion of meaning. It occurs when a word in the target language (which shares at least a meaning with its foreign analogue) takes on a new meaning through imitation. An example is the Italian verb 'realizzare', which means 'to build, to make', but now also 'to understand', under the influence of the English 'to realise'. On the other hand, structural calques are actual *mot-à-mot* translations of foreign expressions. The phenomenon can happen with single terms (e.g. Italian 'fuorilegge' that comes from English 'outlaw') or entire syntactic constructions and idiomatic phrases (e.g. Italian 'non importa cosa' that comes from French 'n'importe quoi').

language, it must be remembered that audiovisual translation is fundamentally different from other types of translation. The reason is because, as already determined at the beginning of this chapter, AVT works with multimedial objects in which a vast array of semiotic modes coexist. Language, which remains one of the main semiotic modes, is accompanied by a constellation of elements (images, sounds, gestures, clothes, light effects, camera movements, songs, to mention but a few) which work together to build meaning and compensate for any possible deficiency. Furthermore, it should be noted that the field of AVT is constantly redesigning itself, and thus the features mentioned above are not fixed and static, but rather subject to evolution and innovation. Pavesi (2005: 35), for instance, notes how contemporary dubbed Italian frequently includes colloquial traits from non-standard varieties, dialectal elements and multilingual mixtures in order to recreate the geographical and sociolinguistic varieties of the original. For what concerns English-language dubbing, a similar point can be made. Hayes (2021) proposes an interesting overview on Netflix's translation choices in the context of non-English language films and TV series dubbed into English. In particular, the author analyses the creative solutions adopted for source texts in Spanish containing instances of language variation. As a matter of fact, Netflix has started to experiment using anglophone variation[37]. Taking advantage of the mostly convention-free English-dubbing industry, the company wanted to provide viewers with multidimensional characterisations similar to those experienced by original-version viewers. This "bold move […] disrupted industry norms for mainstream audiovisual translation […] into English", which "generally strived and strive still […] for standardization" (2021: 2). Although results are not always qualitatively satisfying[38], these attempts can be classified as instances of 'creative' translations. Indeed, as Romero-Fresco and Chaume (2022) puts it:

> "Creativity is a buzzword hailed by academics and professionals as the key to success in the world of media localisation. Creativity is sometimes understood as the best way to honour the original […], but other times it is used to refer to all forms of non-canonical solutions to localisation problems, in other words, as a deviation from standard (mainstream) translation solutions and from current (mainstream) guidelines" (2022: 75).

[37] Of course, Netflix is the commissioner of translations which are undertaken by external AVT services providers. Hayes explains that until 2017 dubs had used standardised American English accents, whereas from 2018 the company opts for more creative solutions in certain circumstances. This does not relate to Spanish originals only. For instance, Italian Netflix original gangster film *Lo Spietato* (De Maria 2019), starring Riccardo Scamarcio, was dubbed in English in a quite unconventional manner: Scamarcio dubbed himself in English, and the rest of the cast was dubbed by Italian dubbing actors speaking English as well, resulting in a dubbed English version with a strong and persistent Italian accent.

[38] From a qualitative point of view, poor performance is not always a cause of experimentation. As Spiteri Miggiani (2021) notices, this can be due to the lack of familiarisation: "the sudden boom of non-English language content on popular streaming platforms has obliged media localisation companies to […] adapt quickly to provide the market with English dubbing lip-synch services." (2021: 138). This means that "many dubbing actors and translators/adaptators may very well be at their first dubbing attempts, and the lack of a well-consolidated point of reference raises the level of difficulty" (2021: 140).

If creativity finds fertile ground in English-language dubbing because, in a sense, it is still in its infancy, the same thing happens in the field of cyberdubbing, as it is unregulated and often made by amateurs with no professional background. In effect, while mainstream AVT practices typically adopt neutral or non-marked varieties (de Higes-Andino 2014), cyberdubbers are known to be experimenters. When translating films presenting multilingualism or language variation, for instance, cyberdubbers often opt for creative solutions. In this direction, a research of particular relevance is Androutsopoulos' (2010) analysis of German dialects representation through parodic cyberdubbings on YouTube. In particular, the audiovisual text types discussed are the so called 'synchros', i.e. popular humoristic dialect-dubbings of film excerpts (or other types of video content). This constitutes an example of amateur dubbers' use of dialects, which is very unusual if compared to mainstream dubbing. Indeed, the world of internet-based cybertranslation is "a clear example of how an artistic and imaginative contribution to the audiovisual text can reach new audiences and give them a new experience" (Romero-Fresco and Chaume 2002: 82).

To conclude, we could state that, even if exceptions exist today, the abovementioned considerations about the features of dubbese are still to be considered valid and are often the main reason for the comparison between dubbing and subtitling.

1.3.5. Dubbing vs. subtitling: a never-ending story?

A conclusive remark is required to complete the systematic comparison between subtitling and dubbing that emerged during this first chapter. It should be clarified that the present work attempts to avoid any bias and does not aim to determine the best method of AVT. In fact, attempting this is as challenging as it is futile: "until recently there was a debate over whether subtitling is better than dubbing […] but this discussion has now been dismissed by scholars for being obsolete, since the reasons for opting for one over the other are varied" (Boisseaux 2019: 49). As has been previously discussed in detail, historical and political reason have contributed to the emergence of a distinct dichotomy between dubbing and subtitling countries. Today, however, this imaginary map seems outdated, especially considering the spread of self-regulated accesses to films and TV series in subscription video-on-demand (SVOD) services (Ghia and Pavesi 2021: 161). Of course, the choice also depends on economic resources. In the film industry, since the costs of dubbing are high, its use is financially beneficial only in potential markets that are large enough to ensure a return on the investment. In this case as well we have seen how bottom-up practices from users and new technological developments make the economic factor less important.

While all this is true, it is when looking at utility-oriented choices that it becomes evident that a direct comparison is not an option. In other words, although personal preferences play an

important role, the choice between the two modalities is in some cases dictated by other parameters. In the important field of accessibility, in effect, specialists situate "the usefulness of subtitling […] for deaf and hard-of-hearing audiences […], in contrast with the usefulness of dubbing for people with poor reading skills, for children and for the blind and visually impaired" (Matamala et al. 2017: 3).

Another crucial difference is the impact of the two modalities on language learning. As outlined in Section 1.2.1., subtitles are useful tools for language learning in both formal and informal contexts. Among many who have dwelt on this theme, Chaume (2003) explains that:

> "Hearing the original and being able to contrast what we are hearing with what we are reading in our own language encourages the learning of foreign languages, particularly English. It is, for instance, often assumed that people living in countries with a strong tradition of subtitling tend to have a better knowledge of English than those living in countries with a preference for dubbing" (2003: 202).

This aspect is, of course, absent for what concerns dubbing, as the input the viewer/hearer receives is in his/her own language. This point should not, however, be overlooked, as dubbing has also been used historically as a vehicle to achieve literacy or linguistic uniformity in a country, as it occurred in France (Pérez-González 2009: 18).

It is not even easy to ascertain which technique offers more advantages from a technical and objective point of view, since both are constrained in some way (i.e. syncrhonisation for dubbing and space-time limitations for subtitling). The discourse on constraints has traditionally caused great debates and stances in research. Furthermore, "beyond the academic arena, many voices have often criticised dubbing as a monster that destroys the artistic quality of the original film […], but the same has also been done for subtitling" (Matamala et al. 2017: 1). The respect for the original text is, at first glance, a relatively straightforward concept. However, a closer examination reveals that it is a more complex matter. Indeed, if "subtitling respects the original voices of the characters, […] dubbing respects the original image" (2017: 3). For this exact reason, researchers sometimes consider dubbing as more immersive and satisfying than subtitling, because the linguistic transfer is integrated into the already multimodal semiotic context of the film (e.g. Ghia and Pavesi 2021, Perego et al. 2015). The features of the language of dubbing outlined above, however, raise questions about the degree of naturalness and contextual appropriateness of such modality (Pérez-González 2014: 50). With regard to dubbese's perceived artificiality, Pavesi (2005) makes an interesting point, namely that different film genres (and different production objectives) typically require different types of translation (2005: 10). From this perspective, it is legit to expect that realistic (dramas or thrillers) and unrealistic films (musicals or animated films) will have different

degrees of closeness to spontaneous speech. In addition to this, for what concerns the perceived naturalness of dubbers' voices, studies have shown that there is no need for them to be similar to original actors' actual voices[39], and that it is only when relevant TV series characters' dubbers are replaced that audiences complain about new voices sounding unnatural and artificial (Herbst 1997: 291-292).

The respect of the original and the degree of naturalness are key factor that will also be addressed later in the discussion of novel forms of AVT with generative artificial intelligence.

[39] Herbst specifies that if the casting is well directed, dubbers' voices will sound as natural as the original ones. If original actors are well known, and so are their voices, the situation might change. This is the reason why, famous actors are (or should be) always revoiced by the same dubber.

CHAPTER 2
AUDIOVISUAL TRANSLATION IN THE AGE OF ARTIFICIAL INTELLIGENCE

2.1. Artificial intelligence revolution

Technological development is heavily modifying the world of AVT. As mentioned on several occasions, technology has today redesigned the relation between AV content and the audience. Indeed, mobile devices, social networks and over-the-top (OTT) platforms[40] are now working together causing an unprecedent global growth in the demand of translated content. Nevertheless, if we categorise as technology all inventions and advancements that have shaped AVT processes and techniques, it becomes clear that the phenomenon is by no means new. As a matter of fact, it would be even difficult to summarise here all the various implementations that have emerged over time. To give an example, as seen in Section 1.3.1., the advent of sound-on-film technology boosted the creation of dubbing as a new AVT method. More recently, the introduction of template files – "the Holy Grail of the subtitling history" (Georgakopoulou 2019: 138) – at the turn of the century was a revolution that dramatically accelerated the subtitling process, as they consisted in files "that contained the source language subtitles already timed and segmented" (O'Hagan 2020: 565), i.e. ready to be translated in other target languages.

Until a few years ago, scholars referred to the digital and hi-tech progress characteristic of the 21st century as the fourth (industrial) revolution (e.g. Schwab 2016; Floridi 2014). Today, with the advent of generative AI, some are talking about a fifth revolution, as this new technology has the potential to radically change most aspects of modern living – including, of course, the audiovisual world.

To be more precise, 'AI' is a general term referred instead to particular sets of algorithms that allow models[41] to perform tasks thanks to machine learning (ML) processes[42]. Generative AI, when applied to language, consists of pipelines[43] of Natural Language Processing (NLP) tasks. On its part, NLP is the branch of computational linguistics that deals with the performance of linguistic tasks, the earliest one being machine translation (MT) in the 1950s (Jezek and Sprugnoli 2023: 17).

[40] 'OTT platform' is another term used to refer to subscription-based streaming platforms such as Amazon Prime Video or Disney+, which provide content on the Internet.

[41] Computational models are aimed to perform the most varied tasks (e.g. face recognition). In order to carry on tasks of linguistic nature (e.g. natural language generation, speech synthesis), today's state-of-the-art models are the large language models (LLMs), such as the groundbreaking Openai's *GPT* (on which the chatbot *ChatGPT* is based) or Google's *Gemini*.

[42] Machine learning is the process through which computational models learn to perform tasks based on the repeated exposure to forms of experience, i.e. enormous amounts of data (Jezek and Sprugnoli 2023: 26).

[43] 'Pipeline' is the term referred to a set of data processing tasks that follow each other in order to perform a more complex task.

From the first rule-based MT to statistical MT, the research has now advanced to neural machine translation (NMT), which is performed through neural networks approaches, today's state-of-the-art in NLP in general. The relevance of this innovation is not to be underestimated: many scholars agree to define MT as "the most profound change yet in the role of the translator" (de los Reyes Lozano and Mejías-Climent 2023: 3), and research shows that today "a shocking amount of the web is machine translated" (e.g. Thompson et al. 2024). Therefore, it is not surprising to see this system implemented in AVT procedures. In addition to NMT, audiovisual translation specialists today have at their fingertips other NLP techniques which can revolutionise their work, notably automatic speech recognition (ASR) and text-to-speech (TTS) technology.

ASR systems use algorithms to to convert spoken audio, represented as waveforms, into written text. For this reason, ASR is also commonly referred to as speech-to-text (STT). Equally to the earliest MT, ASR technology was initially rule-based, and presented problems in terms of speech recognition when voices were characterised by particular variations or accompanied by background noises. As of today, however, performances are generally accurate[44] thanks to ML algorithms and to the vast training datasets of the models.

TTS (or speech synthesis) could be considered as the opposite of the previous task, since it consists of the conversion of written input into audio output. TTS systems as well are today able, thanks to machine learning algorithms, to achieve performance levels that were unimaginable until a few years ago[45]. ASR and TTS are nowadays implemented in all digital assistants, such as Amazon's *Alexa* or Apple's *Siri*, which manage to automate processes involving vocal inputs and outputs (e.g. give directions or make calls) through a combination of NLP techniques. In parallel, such techniques allowed the development of new AVT modes, as in the case of respeaking (Section 1.2.1.1.) and the much less established automatic dubbing (AD[46]).

On their part, big entertainment industries have already seen in this field a significant economic opportunity and have started experimenting to anticipate trends and maximise profits. For instance, an Amazon AI research team developed a new AD model focused on the control of verbosity (i.e. the length of the translation output), with the objective of producing translations that are not only linguistically correct, but also long enough for the automatically dubbed sentences to match the original speech ones (Lakew et al. 2021). While this control is a routine task for human audiovisual translators, it is particularly challenging in the case of AD, and apparently has very

[44] There are various measures to evaluate the performance of ASR systems. Examples are the Word Error Rate (percentage of mistranscribed words), the Sentence Error Rate (percentage of mistrancribed sentences), and the Phoneme Error Rate (percentage of mistranscribed phonemes).

[45] In the case of TTS, performance evaluation measures can be either objective (i.e. the same as those used for ASR) or also subjective, with qualitative judgements of the speech synthesis result. Examples are the Diagnostic Acceptability Measure (where clarity and naturalness are rated from 1 to 5 by listeners) or the Paired Comparison Test (where human listeners choose the sequence that sounds most natural to them between a natural speech sequence and a computer-generated one).

[46] From this point on, the acronym AD will refer exclusively to automatic dubbing, and no longer to audio description.

positive outcomes on viewers (2021: 7541). Bytedance, the Chinese giant owner of TikTok, approached the issue of synchronisation from a different perspective. Its research team presented *Neural Dubber*, the first multimodal AD system which synthesises speech in the target language using the original video's lip movements as a reference to control the prosody in the output (Hu et al. 2021).

Before going into detail and explain what is precisely intended when we talk about AD, it seems appropriate to provide now a brief overview of the issues and risks associated with the use (or misuse) of automatic dubbing and AI in general.

2.1.1. Issues and risks

The pros of this still emerging yet already impactful technology are many. First, AD reduces the need for dubbing actors and much of the traditional studio work of recording and post-production. Cost-effectiveness and time saving, in particular, seem to be the only parameters to guide companies in the new audiovisual landscape, where content proliferates, and localisation demand is higher than ever. Nevertheless, AD has recently gained worldwide attention not only for its potential, but also for its associated risks. This can be applied, as a matter of fact, to artificial intelligence in general, a domain still surrounded by numerous doubts.

A much-discussed example in recent times are deepfakes, i.e. hyper-realistic images, videos and audio created using generative AI. Deepfakes are often produced for entertainment purposes: by exploiting the technologies behind the AD that will be analysed in more detail below, to give an example, users can create music covers that never existed, like famous Donald Trump's version of rapper 50 Cent's song *Many Men*[47], appeared online after the assassination attempt on the politician.

But the technology, which the press is starting to consider as "scary" (Todasco 2023), also lends itself to the creation of fake news (Satariano and Mozur 2023), scams (Thompson 2024), explicit or intimate non-consensual material (Mackenzie and Choi 2024), as well as to the manipulation of public opinion in the field of politics (Appel and Prietzel 2022). Deepfakes are, thus, only the tip of the iceberg.

Narrowing it down to the AVT industry, there is a growing concern about new AI-based tools. Balancing technological advancement with the preservation of human labour is becoming a key challenge. The issues are of different orders. From a qualitative point of view, it is important to assess whether the use of these tools will lead to the demise of creativity or to the underestimation of human performances' nuances. In economic terms, industry stakeholders are attracted by the efficiency of these tools, regardless of the potential displacement of human workers, who are already experiencing difficulties in their roles within the AVT sector.

[47] See https://www.youtube.com/watch?v=f90BL4uVIag&ab_channel=HiRezTV

Indeed, "towards the end of the 2010s, dubbing seems to have entered its golden age" (Banos 2023: 62), mainly due to the explosion of non-English-language shows and films on OTT platforms. Nevertheless, the moment does not seem to be so 'golden' for AVT professionals. The reason is because, as many sector experts declare, the insane volume of work is not commensurate with the pay that companies offer to professionals (Bryant 2021). Romero-Fresco (2013: 201) observes how, paradoxically, more than half of the revenue of most big-budget Hollywood films comes from their accessible or translated versions for foreign markets, while only 0.01 to 0.1% of the budget is spent on the translation process. If professionals are poorly paid, the quality of their work could be affected. In this scenario, new AI tools can become serious competitors to human translators, being imposed or integrated by companies to accelerate processes and match the high demand.

Some scholars believe that fully automatic dubbings are likely to become the norm in AVT future (O'Hagan 2020: 567). If this cannot be entirely confirmed yet, it is however likely that human intervention will be reduced more and more, and probably be confined to the manipulation of the input and the output texts. AV translators' job could become, in other words, a polishing and PE (post-editing) job, where "PE means to review a pre-translated text generated by an MT engine against an original source text, correcting possible errors to comply with specific quality criteria" (Guerberof-Arenas 2019: 334). Indeed, "the very term translator is [already] being questioned […] in favour of new terms such as post-editor and text localizer" (de los Reyes Lozano and Mejías-Climent 2023: 2). In the same vein, studies have been published to underline the necessity to include innovative AVT methods in the training of future translators, because "the impact of technology must be reflected in the education strategies to meet the requirements of a rapidly evolving market and equip students with the necessary knowledge and skills for professional activity in the digital era" (Wang 2023: 1526).

As aptly reminded by de los Reyes Lozano and Mejías-Climent (2023):

> "In view of the many concerns that AI […] generate among AVT professionals and audiences, putting on a blindfold and acting as if technology did not exist might not seem to be the best approach to the issue […]. On the contrary, defending labour rights, improving the working conditions of practitioners, and respecting certain ethical values is essential" (2023: 4).

2023 was the year of the record-breaking strike of Hollywood actors and screenwriters. In particular, the SAG-AFTRA (Screen Actors Guild – American Federation of Television and Radio Artists) managed to secure a historic deal from the studies that included, among its most important points, new rules on the indiscriminate use of AI (Scherer 2024). This agreement sets a precedent

in AI history, and constitutes a resounding example, yet a number of other initiatives have emerged with the objective of regulating this new technology.

In April 2024, for instance, the European Federation of Audiovisual Translators (AVTE), released a document, the *Statement on AI regulation*[48], which is aligned with other publications by analogous creative authors' associations around the world, and raises important questions on four different points. First, the statement focuses on issues related to technical aspects of AI, with one of the main concerns being that "generative AI doesn't come up with original concepts by itself […], [as] it re-elaborates existing (possibly copyrighted) materials, making artistic works less and less original […] [and going] against the best interests of the original authors" (AVTE 2024). The second point regards the industry, and mainly focuses on the Federation's strong disapproval of the devaluation of the translator's role, which is often poorly paid and relegated to tasks of machine translation post-editing (MTPE) or "light MTPE, where translators are instructed to just fix blatant errors instead of producing quality output […] and audiences are presented with more and more mediocre texts" (AVTE 2024). In this direction, we could mention an observation by Romero-Fresco and Chaume (2022), who view human touch as indispensable in the field of AVT both for the respect of the creative intent and for the vindication of the translator's visibility (2022: 77). Finally, the third and fourth points of the statement address the issue from a legal and an ethical perspective, focusing on themes such as authors' rights, human creativity, cultural homogenization, and working conditions.

Copyright protection has become a theme of utmost importance in recent years, after the explosion of generative artificial intelligence models, which are trained on large amount of data "collected from various resources such as the Internet, even without the permission of the original data owner" (Ren et al. 2024: 2). Ren et al. (2024) tackle this issue considering different types of "Deep Generative Models (DGMs)" (2024: 2), i.e. generative AI models able to create synthesised content including text, images, and audio. In this context, the authors identify three parties involved in the discussion. First, there are the "Source Data Owners", i.e. the producers of the data, those who actually possess the copyright of the source data used for the training. Secondly, there are the "DGM Users" and the "DGM Providers", who in both cases "have reasons to demand the copyright of the generative contents" (2024: 2), as the output generated through the models are not only obtained thanks to the training data, but also thanks to users' creative prompts and providers' efforts in the engineering of the model. Although the authors underlines that this is still "a complex and evolving legal and ethical issue" (2024: 2), technical solutions to the problem are proposed for each of the parties involved. For what concerns the object of study of the present thesis, i.e. synthesised speech, many are the contributions that explore the risks in terms of

[48] See https://avteurope.eu/2024/04/29/avte-statement-on-generative-ai/.

copyright that this technology entails. An example is provided by Zhang et al. (2024), who provide a protection technology to prevent the generation of unauthorised high-quality deepfake speeches based on publicly accessible speech data that contains sensitive information.

Rapid changes in society over the last years have indeed highlighted the necessity for policymakers to address the subject in a serious manner. In this new context, regulating AI is vital for at least two reasons: first, legal frameworks lead people to understand the world around them; second, if there is a solid legislation, companies, citizens, and states can design projects following shared directions and intentions, without being overwhelmed by the progress.

In 2018, a scientific committee chaired by philosopher Luciano Floridi published *AI4people – an ethical framework for a good AI society*. The paper presents a series of recommendations for the responsible and ethical development of AI, taking into account both its potential benefits and perceived risks. As the authors points out, "we can safely dispense with the question of *whether* AI will have an impact; the pertinent questions now are *by whom, how*, *where*, and *when* this positive or negative impact will be felt" (2018: 671). For instance, the idea that AI could enable humans to free themselves from superfluous labour as many inventions in the past have done is questioned by the pace at which the spread of such technologies is happening, which may cause "a very fast devaluation of old skills and hence a quick disruption of the job market" (2018: 673). In this perspective, job losses are not a direct consequence of the technological innovation, but the result of an unbalanced relationship between demand and supply. Overall, the document highlights how "AI can be used to foster human nature and its potentialities, thus creating opportunities; [but also] underused […] or overused and misused" (2018: 690). If overuses or misuses are easier to identify, AI underuse is an interesting concept: "fear, ignorance, misplaced concerns or excessive reaction may lead a society to underuse AI technologies below their full potential, for what might be broadly described as the wrong reasons […]. As a result, the benefits offered by AI technologies may not be fully realised by society" (2018: 691).

The document (written at a time when institutions had not yet taken a position on the issue) was presented to the European Parliament. Later, a commission for AI was created, along with the *AI Act*, "the first-ever comprehensive legal framework on AI worldwide"[49]. Proposed in 2021 and entered into force in August 2024, the Act identifies AI applications and ranks them by their risk of causing harm. There are four levels (minimal, limited, high, unacceptable) and each level is related to specific precautions and directives. For example, AI applications that allow the generation of synthetic images or sounds (like deepfakes, or AD software), fall into the 'limited risk' category, and for this reason require transparency obligations, i.e. users need to be informed about the nature of the system or of the product they create or consume. On the other hand,

[49] See https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai .

applications that are considered unacceptable (which include, for instance, systems for the real-time remote biometric identification in public spaces) are entirely banned. The European legislation was matched by the *Executive Order 14110*[50], which is considered to be the most comprehensive legal framework about AI by the United States government. Published in October 2023, it also provides directions and stresses the need for a "responsible AI use […] to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure"[51].

From a legal perspective, sudden developments of novel technologies make total regulations difficult. However, even if innovation in technology and law have different paces, frameworks are still needed for ethical and sustainable futures. From an academic perspective, this "black mirror effect"[52] can be tackled by different areas of study, including AVT studies.

2.2. Automatic dubbing

Automatic dubbing consists of the process of replacing a human voice present in an audiovisual text with a synthesised voice in a different target language, "while preserving as much as possible the user experience of the original video" (Lakew et al. 2021: 7538). This last point is of particular importance, because, as Federico et al. (2020) explain, it is what distinguishes AD from speech-to-speech translation:

> "Automatic dubbing can be regarded as an extension of the speech-to-speech translation (STST) task […], which is generally seen as the combination of three sub-tasks: (i) transcribing speech to text in a source language (ASR), (ii) translating text from a source to a target language (MT) and (iii) generating speech from text in a target language (TTS). […] The main goal of STST is producing an output that reflects the linguistic content of the original sentence. On the other hand, automatic dubbing aims to replace all speech contained in a video document with speech in a different language, so that the result sounds and looks as natural as the original. Hence, in addition to conveying the same content of the original utterance, dubbing should also match the original timbre, emotion, duration, prosody, background noise, and reverberation" (2020: 257).

[50] Full title: *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*.

[51] See https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[52] The black mirror effect is "the phenomenon of recognizing the possible outcomes of rapid technological advancement, but also to encourage critical thinking about how we integrate technology into our lives and interact with it" (de los Reyes Lozano and Mejías-Climent 2023: 2), The expression 'black mirror', popularised by the famous TV series *Black Mirror* (Brooker 2011-), refers to the screen of a mobile device when it is switched off, which allows us to see the reflection of ourselves.

From this workflow explanation we understand, once again, that when AD is referred to as 'AI (generated) dubbing', the label 'AI' is in fact used as an umbrella term covering a wider set of machine-learning-based tasks. Automatic dubbing is also sometimes referred to as machine dubbing, or video dubbing, since, as Wu et al. (2023) explain, its "cascaded system" is particularly difficult to master due to the constraints that are inherently built in the videos (i.e. length and duration of the translated speech must match the video). However, in the present study, the term AD will be employed.

A cursory survey on search engines like Google reveals a multitude of websites of companies that provide end-to-end AD services. In most cases, the service provided is defined as AI dubbing, although the line separating it from STST is very thin. This is because at the moment, AD works better in fields which do not require lip synchronisation (as most of the speech comes from off-screen voices), and where traditional dubbing would be unnecessary or not convenient. "Daily news, for instance, cannot be dubbed into other languages by traditional methods because of its quick turnaround times" (Team Papercup 2024). The company Papercup provides an example of successful AD for news broadcasting. Indeed, Sky News automated news localisation on its YouTube channel *Sky News en Español*[53], where videos are entirely revoiced from English into Spanish using Papercup technology. Educational videos are another audiovisual textual genre prone to AD. Baños (2023) analysed the naturalness and accuracy in the videos uploaded by the YouTube channel *Amoeba Sisters en Español*[54], automatically dubbed into Spanish from the original English videos in the main channel[55], using the free tool Aloud (developed by Google). Even though the results still show much room for improvement, these examples constitute legitimate attempts for the global diffusion of original and accessible content. On its official blog, YouTube claimed to be working in the direction of a responsible regulation of content that is identified as altered or synthetic (O'Connor and Moxley 2023). This is particularly significant because, in the context of online platforms where speed is critical to maintaining competitive advantage, new technologies can be overused. Indeed, by reducing both processing time and associated costs, AD offers social network users the possibility to accelerate workflows and boost the production of audiovisual content. However, this acceleration often leads to huge amounts of copied non-original AI-revoiced videos, which are published for monetisation and raise questions about authorship and creativity.

[53] See https://www.youtube.com/@skynewsespanol
[54] See https://www.youtube.com/@AmoebaSistersEspanol
[55] See https://www.youtube.com/@AmoebaSisters

2.2.1. Understanding synthesised speech

With regard to the last step of the AD workflow, the speech synthesis, it is worth mentioning that significant advancements have been made in recent years.

It may be now useful to underline that the main concern in AD research seems to be related to the TTS stage. In particular, compared to human dubbing, AD appears to struggle in achieving isochrony, i.e. "when the dubbed speech perfectly matches the timing of the original utterances and pauses" (Effendi et al. 2022: 8037). "The critical issue with these [new and automatic] approaches is that speech generated via TTS in the target language have a different length than the corresponding segment in the source language […]. Hence the dubbing output look unnatural from an audio-visual synchronization perspective" (Sahipjohn et al. 2024: 1).

Indeed, many experiments have been recently conducted to control the speech duration of synthesised speech in a way that it aligns well with the speaker's lip movements, as in the case of the already mentioned study by Lakew et al. (2021) about verbosity control, or of Saboo and Baumann (2019), who integrated dubbing constraints into the machine translation step. Although this constitutes an essential point for the success of automatically dubbed versions of audiovisual products – and will be touched later in the reception study – other aspects deserve attention.

For example, other important criteria for the determination of the quality of an AI-generated dub is voice naturalness. Although the term 'naturalness' can be vague and open to many interpretations, this is here to be intended as the ensemble of elements that makes the synthesised voice in the output as realistic and as close as possible to the original (e.g. pronunciation, vocal timbre, prosody, emotional expression).

Whereas most systems were already capable of generating almost lifelike synthetic voices, tools are nowadays powerful enough to go beyond this. Machine learning algorithms allow models to analyse training voice samples, discovering patterns and extrapolating features to ultimately replicate an existing voice. Indeed, companies are starting to experiment with the speaker-adaptive text to speech technology, which enables to entirely clone the voice and emotional tone of the source speaker. This is the case of the system developed by DeepMind (a company acquired by Google), which succeeds in the task of speech synthesis with voice imitation. The development team (Yang et al. 2020) explains that the entire process is quite long and laborious, as pre- and post-processing tasks still need to be taken care of manually (e.g. errors in the translated transcript have to be corrected and utterances in the output need to be aligned). Voice imitation is obtained thanks to a fine-tuning of the TTS model used for the processing of the output. This means that the model, which is already "trained on a large multilingual multi-speaker dataset", is further enriched with speech data from the original video (i.e. the source video that needs to be dubbed) to achieve "good quality in naturalness and voice similarity" (2020: 15).

It is evident how voice reproduction can significantly enhance the quality of an automatic dub. In addition to this, capturing and reproducing the source actors' emotional nuances in the delivery of the lines is what really distinguishes AD from STST, and what could allow the technology to be implemented in professional and marketable contexts. Chen et al. (2022) published a study that proposes to extend the pre-existing tasks of Voice Cloning (VC) (which "aim to convert a paragraph text to a speech with desired voice specified by a reference audio"), through a new task named Visual Voice Cloning (V2C), which "expects a resulting speech with the same voice but varying emotions derived from the reference video" (2022: 21210). Voice (and emotional expression) imitation in synthesised speech could be of pivotal importance and open to new future opportunities, for both the audience and the industry. On one side, viewers could have a totally new and immersive experience. On the other, distributors could take advantage of AD for localising entertainment content such as films and TV series (e.g. Genelza 2024). Even though much information is not to be found online, this technique has been already experimented by the streaming platform Hulu, which partnered with the company Deepdub to revoice the Portuguese show *Vanda* (Müller 2022-), the first AI dubbed drama[56]. Although human intervention was still required (Welk and Maglio 2024), many professionals see this as a slippery slope that could cost many jobs (Fuster 2024), as enhanced voice cloning in the future could make distinguishing between human and computer-generated voices almost an impossible task.

On the other hand, there are companies which offer alternatives to enhance the dubbing process without the objective of removing human professionals from recording booths. It is the case of Flawless, which has developed the new *TrueSync* technology. The software in question does not generate voices, but is able to digitally modify the video output, so that actors' lip movements are recreated from scratch and match the translated speech in the desired target language[57][58]. *TrueSync* is already commercially available and featured among the Time's "best inventions of 2021"[59].

[56] See https://deepdub.ai/

[57] This is useful both for solving the old problem of synchronisation in interlingual dubbings and for avoiding the reshooting of imperfect scenes.

[58] *TrueSync* is a recent technology, but many attempts have been made to master the manipulation of facial movements from a source to a target video. A similar example was introduced by Thies et al. in 2016. The work presented a new approached called *Face2Face*, "the first real-time facial reenactement system that requires just monocular RBG input" (2016: 2393). In practice, starting from any video containing a facial performance, the facial expressions of a source actor were captured by a webcam and animated in real time on the face of a target actor in the target video. By reproducing not only the expressions, but also the inside of the target actors' mouth, this innovative technology aimed to open up new avenues in video conferencing as well as in interlingual dubbing. Another interesting contribution in this field is provided by Kim et al. (2019) research, which present a style-preserving visual dubbing approach. This approach was developed with the desire of creating automatic dubs modifying facial expressions while maintaining the identity-specific idiosyncrasies of the source actors.

[59] See https://time.com/collection/best-inventions-2021/6112554/flawless-ai-truesync/

Indeed, although not all scholars agree on the importance of synchronisation[60] (Pavesi 2005: 16), this is undoubtedly a ground-breaking introduction that turns the tables of AVT. Without considering Hollywood's huge production machines, however, the creation of automatically dubbed content has increased even on the user side. As it will be described below, social networks are the privileged place for the proliferation of such content.

2.3. New cyberdubbing cultures

This section aims to provide an insight into the cyberdubbing Italian landscape. To do so, the first step is to identify, by searching on various social networks, examples of pages, channels, or content creators that correspond to what is being examined here, i.e. the new AI-based cyberdubbing practices.

Social networks are largely known to be characterised by democratisation, enabling interpersonal communication, (almost always) free access to content, and the possibility to spread creative productions as well. Users living these cyber-contexts are not always aware of the legal status of their activities. Translated content poses questions about legitimacy, as

> "in the media convergence, translation can exist in a space that shifts constantly between the legal and illegal slides of the media industry […]. Non-professional translation operates in grey areas that are constantly redefining the roles of translators […]. Audiences do not necessarily need to choose a side; they navigate the different legal and illegal systems to have all their needs covered" (Orrego-Carmona 2018: 336).

Dwyer (2019) mentions how the new fan (yet this is valid for all cyber-) AVT forms, which are "largely amateur, free, unregulated and even illegal" (2019: 436) are left out of the academic and industry discourse. However, their growth "registers another home truth about AVT as a whole: it is hostage to the winds of technological change" (2019: 436). New forms of translation production (including AI-generated ones) deserve thus equal dignity in research terms. Indeed, as explained by Pérez-González (2014):

> "following the advent of collaborative technologies, the study of audiovisual translation can no longer be based on the premise that the field is exclusively in the hands of professional mediators. The boundaries that have traditionally prevented media consumers (in this case, viewers) from taking on the role of amateur co-creators, either by collaborating with other

[60] The intensity of the synchronisation constraint can vary according to the type of scene and the type of film. Moreover, interest in this aspect generally varies from country to country. In Italy, for example, dubbing is a well-established habit and the audience is traditionally more demanding.

professional agencies throughout the life- cycle of a project or by modifying what professionals have previously produced, are no longer enforceable" (2014: 66).

For the purpose of this study, cyberdubbing examples have been searched on the photo and video sharing social network Instagram. A review limited to Italian Instagram accounts only reveals numerous content creators engaged in the practice of AI-powered parodic cyberdubbing, not to mention the re-shared and viral material on the platform, the origin and ownership of which is sometimes hard to establish. As can be seen from *Table 2* (which outlines the characteristics of some illustrative accounts), the type of content may vary, yet the objective always appears to be the achievement of some humorous (or, at least, entertaining) effect.

Table 2: Overview of some Italian Instagram accounts publishing automatically dubbed content

Instagram account	Followers	Main content
speechif.ai[61]	40,300	Serious interlingual AD (other languages > ita) of famous people's interviews
sandrumasi[62]	21,200	Parodic interlingual AD (ita > other languages), parodic intralingual AD
italiancomedydub[63]	32,000	Parodic interlingual AD (ita > eng) of clips from famous Italian comedy films

The three Instagram profiles mentioned above were contacted privately with the aim of providing this discussion with more precise details on the content creation procedures. In order to gather information about the projects and the technical processes involved behind the publishing of such content on Instagram, an informative questionnaire (see Appendix A) with open-ended questions was created[64].

Although dubbing into Italian and not from Italian (which is the main focus of this study), the account 'speechif.ai' is nevertheless interesting to mention for at least two reasons.

First, the videos posted in this page are not just an example of AD, but they are also instances of the technology that captures original actor's face features and generates new lip movements that match the new language spoken in the target video. Although this result is similar to what described

[61] See https://www.instagram.com/speechif.ai/
[62] See https://www.instagram.com/sandrumasi/
[63] See https://www.instagram.com/italiancomedydub/
[64] The questionnaire is in Italian, since the selected accounts all belong to Italian native speakers.

in the previous section when the *TrueSync* technology was described, the account's admins were not available to answer to the informative questionnaire regarding the procedures used for the creation of their content, so the actual technology used in this case is unknown.

The second interesting aspect concerns the videos' comment sections. As specified in *Table 2*, the primary content of the page is constituted by interviews that have been dubbed in Italian. In each video's caption, the account's admins always include a notice[65] warning users of the artificiality of the dub. However, even a rapid examination of the comments under the videos posted by the page reveals two interesting aspects. First, it is evident that a number of commenters are unaware that the video has been dubbed using AI tools (and some appear to be unaware that the video has been dubbed at all). *Figure 1* illustrates two examples of comments generated in response to a video interview[66] with a Turkish Olympic athlete that was dubbed into Italian using AI.

Figure 1: Examples of comments referring to the language spoken in the video

Second, another significant portion of the commenters seem not to be focused on the medium (i.e. the fact that the video presents a synthesised speech which clones the voices of the people portrayed), but rather on the content of the video, i.e. the statements and actions of the celebrities in question, as well as any statement or action previously attributed to them. *Figure 2* show two examples of comments under another video[67] posted by the page in the context of the Paris 2024 Olympics, which consisted of an interview with an Algerian boxer at the centre of a heated debate after a match against an Italian boxer.

[65] Notice in Italian: "Parole, voce e labiale sono stati tradotti con l'intelligenza artificiale". English: "Words, voice and lip movements are translated using artificial intelligence".
[66] See https://www.instagram.com/p/C-PdKRUMDM7/
[67] See https://www.instagram.com/p/C-NHdPTMG5U/

Figure 2: Examples of comments referring to the content of the video

The presence of comments of this kind could be indicative of two potential scenarios (which are not necessarily mutually exclusive): on one side, they could be an index of users' increased degree of tolerance towards AD; on the other, they could mean that AD quality is improving to the point that it is now being integrated into the existing AVT methods as a legitimate alternative. In this perspective, AD simply becomes a means for the dissemination of translated material, and users' attention is not drawn by the AVT method as entertaining on its own.

In contrast to 'speechif.ai', the account named 'sandrumasi' (also briefly described in *Table 2*) accepted to provide answers to the informative questionnaire. In this case, the responses were also interesting for a variety of reasons. First, the Instagram page in question is individual and personal, i.e. only one person is in charge of creating the content that is published online. This aspect underlines the essentially free nature (since there are no intermediaries and directives from above to follow) of the online cyberdubbing practices, especially if aimed at creating ironic memes. The other noteworthy aspect is that the person involved in the creation of this content do not study or work either in the field of translation and dubbing, or in computer science and artificial intelligence. This answer highlights the initiative of online amateurs, but also the availability of means that allow anyone to express their creativity and even gain a certain audience. For what concerns the technological means involved, the account's admin indicated two software (Elevenlabs, Rask.ai) for the generation of automatic dubbings. However, the technical questions evidenced how, in practice, dubs are not fully automatic. On the contrary, human intervention is needed in the following passages: translation (software present MT tools, but the output is not always accurate), audio-video synchronisation (even though in some cases, similarly to what 'speechif.ai' does, software incorporating the lip movement recreation technology are used), and emotional tone management (e.g. adding exclamation marks to the text to convey more emphasis to the sentence, or writing in capital letters to obtain a more shouted output). With regard to the speech synthesis,

the admin stated that he makes use of the text-to-speech tool by training the software with excerpts of audio containing the voices that will be then cloned and synthesised. In response to the question of who or what has inspired him for the creation of his content, the admin mentioned the 'italiancomedydub' page as a significant influence.

As it can be observed in *Table 2*, 'italiancomedydub' is based on the production of English-dubbed clips from famous Italian comedies. The affinity between English-language dubbing and cyberdubbing has already been discussed in Section 1.3.4., in terms of the freedom afforded by the lack of rooted conventions. Since cyberdubbing and English-language dubbing are key points of our research, the page in question is of particular interest to us. In this case too, admins were contacted and accepted to respond to the informative questionnaire.

The project was launched in 2023, initially on TikTok. In parallel, a page was open on Instagram, which is currently the platform where the project has collected the most followers (see *Table 2*) and the most views (more than two million in August 2024). The first interesting aspect to highlight regards the answer to the question "Qual è lo scopo del vostro progetto?"[68]. Indeed, the admins stated that "lo scopo principale è la divulgazione di clip comiche italiane a un pubblico internazionale"[69], but also that their content can be "un ottimo strumento per imparare la lingua inglese"[70]. This answer is interesting for two main reasons. Firstly, it confirms the positive attitude towards dubbing by Italians, who are traditionally accustomed to dubbing as an AVT method. Since dubbings are done in English, however, the attitudes of English native speakers towards such content must be taken into account. For this reason, this aspect will be the main object of research in the reception study that will be conducted later. In the second place, since this content is considered by its creators as a helpful tool for language learning, it is worth noticing that, in this way, videos are not only directed towards an international audience, but also towards a national one. Indeed, admins declare that 90% of their audience is Italian. The correlation between familiar and popular films and their dubbing into another language for the purpose of learning the latter will not be examined here, but can constitute an interesting aspect in future research.

In the case of 'italiancomedydub' too, the two individuals engaged in the realisation of the content lack both study and professional experience in the fields of translation, dubbing, and artificial intelligence. For this reason, the same considerations made above regarding the nature of cyberdubbing apply here. One innovative aspect worth mentioning, however, is the collaborative nature of the project. Indeed, the admins of the page declare that they use the direct support of followers for the creation of translations. In particular, the collaboration takes place on an external channel, the instant messaging service Telegram, where the admins share the Italian transcript of

[68] English: "What is the purpose of your project?"
[69] English: "The main purpose is the popularisation of Italian comedy clips to an international audience".
[70] English: "A great tool for English language learning".

the clips. Successively, followers (whose qualification or competence was not specified) engage in the creation of Italian-to-English translations, establishing an informal working relationship and a sort of fan chain. As a matter of fact, if the page's admins publish dubs of famous Italian comedies, they also – in turn – rely on their fans to create translations. This phenomenon takes place in the context of social networks, which have created new types of audiences. As Orrego-Carmona (2018) explains, today

> "viewer engagement goes beyond the active consumption of the content. Active and occasional viewers become involved in the production and consumption of user-generated content that is developed to deepen and enlarge the fictional universes in which the audiovisual products unfold […]. Content [now] can always be expanded, reshaped, amended and corrected, by producers (through remakes, revivals, cross-overs, accompanying websites, social media presence, etc.) or by any of the forms of distribution by users or the creation of user-generated content. This continuum of the creation process and constant expansion of the products through different media give (audiovisual) texts an aspect of non-finiteness" (2018: 324).

In contrast to cyberdubbing, the collaborative nature of cybersubtitling has been largely studied in academic contexts, as it is an established activity on the web. When talking about fasubbers, for instance, many scholars have pointed out the new role of the online user. O'Hagan (2009) talks about user-generated translation (UGT) to refer to a "wide range of Translation, carried out based on free user participation in digital media spaces […] undertaken by unspecified self-selecteed individuals" (2009: 97). This seems to fit particularly well also in the case of cyberdubbing in general, and in the case of the 'italiancomedydub' project in particular. Indeed, "the user in UGT […] is somebody who voluntarily acts as a remediator of linguistically inaccessible products and direct producer of Translation on the basis of their knowledge of the given language as well as that of particular media content or genre, spurred by their substantial interest in the topic" (2009: 97). Similarly, Baños and Díaz-Cintas (2023) emphasise the role of audiences as active participants that are "socially networked and deeply engaged in digital media consumption, production and distribution" (134). In this context, as Gambier (2018) specifies, "fansubbers [„„] are blurring the lines between consumers, users and fans, becoming 'prosumers', […] both using and creating the content online" (2018: 53). These considerations can be therefore expanded, as what mentioned appears to be equally valid for the new cyberdubbing cultures. As regards fandubs, Chaume (2012) notes that "results are far from professional […] but they are not intended to be professional" (2012: 4). The same consideration can be applied to home-made AI-generated cyberdubs which are disseminated on social networks. Thus, the overall quality must always be related to the essentially amateur nature of the products.

With regard to the technical process, the answers show a certain similarity to those provided by the other page mentioned above. In effect, human intervention is primarily present at the translation phase. As already noted, the translation is manual and commissioned to the followers of the project. As it will be discussed in the results of the reception study, the importance of translation stage is not to be underestimated, since its quality can have a significant impact on the overall enjoyment of the dubbed text. This consideration is especially true for the comedy genre, where human translators are necessarily required to check if some particular elements make sense and work in the target language. Typical examples in the field of comedy films and TV series are jokes and puns, highly sociopragmatic elements which are often very complex to render in another language and culture. Let us see an example from a scene in the popular sitcom *The Big Bang Theory* (Lorre and Prady 2007-2019), where the characters are playing Pictionary (a game where players try to identify words based on their teammates' drawings). In the original English version [9], the word to be guessed is 'polish': while the nerd Sheldon (S) draws very specific references to Polish nationality, Penny (P) draws a hand, to denote 'nail polish'. The teammate Amy (A) guesses on the basis of Penny's drawing and this creates the comic misunderstanding. In the Italian version [10], the pun is creatively rendered: instead of 'polish', the word to be guessed is 'lacca'[71], so as to create the misunderstanding with the Polish nationality, which in Italian is spelled 'polacca'.

[9] *Original English version:*
A: "Fingers, nails… polish?"
P. "Yeah!"
S: "No! The word is Polish! See, look, Polish sausage! The model of the solar system developed by Nicolaus Copernicus, a Polish astronomer! […]"
P: "Excuse me, the word is 'polish', see? Small 'p'".

[10] *Italian version:*
A: "Mano, unghia, smalto… lacca?"
P: "Sì!"
S: "No, no! La parola è 'Po-lacca' Ecco guarda, salsiccia polacca, e il modello di sistema solare sviluppato da Niccolò Copernico, astronomo di nazionalità polacca! […]"
P: "Scusami genio, la parola è 'lacca', visto? Hai letto male".

This and many other examples could be used to justify the fact that a fully automatic dubbing – especially in the comedy genre – is very difficult to achieve, as the moments that require complex interlinguistic and intercultural adaptations are many. It is therefore comprehensible that the

[71] English: 'varnish'.

translation stage is handled manually, precisely because of the nature of the scenes that are translated. In addition to this, another reason – however of a technical kind – for the manual translation is that, as already mentioned before,

> "for some pairs of languages, a machine translation model may output translated text with a very different length from the source sentence. This causes problems when aligning the synthesized translated audio back to the original video. Therefore, it is common in the dubbing industry to request a human translator to produce translated sentences with as close length as possible to the original ones. This is still an open challenge for automatic machine translation in the application of dubbing" (Yang et al. 2020: 15).

In any case, translation is not the only manual step in the 'italiancomedydub' content creation workflow. Indeed, the admins indicate that audio-video synchronisation is not automatic but carried out on separate audio-video editing software. The last stage, i.e. the speech synthesis, is particularly interesting to analyse. This is, in effect, the process realised through AI tools. The admins did not reveal which software are used for the task, however they stated that new tools are constantly tested to improve the final product. This aspect denotes the rapidity with which many of these software become obsolete in favour of others, as well as the overall ever changing technological advancements. The speech synthesis process is not, however, entirely automated: the admins declared that: "nel 90% dei casi occorre effettuare un fine-tuning lato IA per cercare di ottenere una recitazione il più vicina possible all'originale, infatti arriviamo a produrre anche 40-60 o più generazioni per ogni singola battuta della scena finché il risultato non è soddisfacente"[72]. This means that, to achieve an emotional tone and an expressivity which is close to the original, all the attempts that are made for a scene that may only span a few minutes, necessitate several hours of work. The admins also denounce a certain difficulty in working on scenes with multiple speakers and overlapping turns ("si prediligono clip con dialoghi chiari e non sovrapposti"[73]) which are, though, normally to be found in films or TV series – because on of the primary objective in fiction is often to represent reality and the spontaneous communicative situations within it. This highlights a clear and evident technical limit in the realisation of these parodic dubbings. Although the large ampunt of work required for the realisation of the clips and the technical limitations included in the process are interesting aspects, they will not be subjected to detailed analysis in the present study, due to the limitations of the research design.

[72] English: "in 90 per cent of cases, fine-tuning is necessary to achieve an output which is the closest possible to the original acting. Indeed, we produce 40 to 60 or more generations for every single line of the scene until the result is adequate".

[73] English: "clips containing clear and non-overlapping dialogue are preferred".

The final questions of the questionnaire concern attitudes towards AI, also from an ethical standpoint. When asked "Siete consapevoli dei rischi legati all'utilizzo di intelligenze artificiali (ad es. deepfake, perdita di posti di lavoro)?"[74], the admins replied to be aware of the risks associated with AI and to believe that it could nonetheless be an opportunity for the emergence of new jobs. Finally, when asked "Pensate/sperate che il doppiaggio automatico possa diventare una realtà nel campo dell'intrattenimento?"[75], they declared to be certain that this will happen, even though not in the short or medium term.

2.4. Research questions

The present work has been developed in light of what has been discussed so far. It would be futile to ignore that "AI applications in translation will continue to put pressure on human translation productivity and cost" (O'Hagan 2020: 568). It is therefore preferable to address the issue, both in terms of regulation and academic research. Indeed, considering the risks associated with possible future scenarios in the use of AI, many scholars think there is a need for "urgent review by academics and professionals" (de los Reyes Lozano and Mejías-Climent 2023: 2). While much has already been written about the incorporation of new technologies in subtitling, "perhaps due to its written mode and technological nature (dedicated software is always required)" (de los Reyes 2023: 4), "the automation of revoicing modes has received less attention in academia" (Baños 2023: 62). Brannon et al. (2023), for example, provide an overview of some recent studies and research about AD, but note a "stark lack of literature on automatic translation of dialogues, compared to common domains in literature like news" (2023: 430). The present work wants to intercept this exact gap in research, since much room for development seems to be available. In particular, we would like to shed light on the phenomenon of amateur AD of fiction, which seems to be underinvestigated yet.

Indeed, in addition to the use of AI for the dubbing of artistic products such as films and TV series, two further themes contribute to enrich the discourse and to outline the research question of this study. These two themes are English-language dubbing and cyberdubbing: similarly to AI, both are rooted in contemporary media landscape and necessitate revision.

Considering the exponential growth in the production of audiovisual content, and thus the increasing demand for translated material fostered by OTT platforms, English-language dubbing has established itself in recent years as a new and tangible phenomenon. As we have seen in previous sections, cyberdubbing shares with English-language dubbing the property of being

[74] English: "Are you aware of he risks involved in the use of artificial intelligence (e.g. deepfakes, job losses)?"
[75] English: "Do you believe or hope that automatic dubbing has the potential of becoming a reality in the field of entertainment?"

conventions-free, mainly because its presence on social networks make it « often deregulated and free from commercial imperatives » (Baños and Díaz-Cintas 2023: 133).

If we pair these two areas of research with new AI-based AVT methods, it is reasonable to believe that traditional boundaries in dubbing are being questioned. As a matter of fact, the development of technologies that allow anyone to translate audiovisual content could have major impacts on both working practices and viewer habits. In this new context, there are many aspects that deserve attention. Narrowing the field, this study aims to examine, with a qualitative approach, multiple aspects related to (English-language) cyberdubbing generated with AI-tools. To do so, the following research questions will be addressed.

RQ 1: What is the reception of native English-speaking audiences to AI-based English-language dubbing?

RQ 2: How do cyberdubbing practices differ from those of professional mainstream dubbing?

RQ 3: What are the opinions and attitudes of native English speakers towards the presence of the new AI-powered tools in the field of audiovisual translation?

In order to answer to this question, a small-scale reception study will be conducted. Indeed, studying the responses of real audiences to real products could resolve these doubts and even broaden the discussion on other topics, such as the potential use of AD in fields like entertainment, and the impact of this technology on the industry. The following chapter will analyse the design of the interview as well as the results obtained. Prior to that, however, the concept of reception will be clarified.

CHAPTER 3
A RECEPTION STUDY

3.1. Studying audiences

3.1.1. Perception vs. reception

Research in AVT should not neglect accurate analyses of the audience and its reception of translated texts. Indeed, "all interaction between a text and its receiver is directed towards a response from the latter" (Chaume 2007: 71). As we highlighted in the previous sections, audience role in influencing the choices of industries and companies is becoming increasingly significant. Therefore, research on audience reception are more important than ever for an understanding of new tastes, preferences and trends in the contemporary media landscape. At this point, it appears necessary to provide a more detailed definition of reception.

In academic and non-academic contexts, the terms 'perception' and 'reception' are often used interchangeably. Nevertheless, substantial differences exist between the two concepts, and the resulting studies. Since we are interested in looking at the reception of translated audiovisual texts, it may be useful to address the dichotomy indirectly, through the overview of research methods employed in AVT studies provided by Pérez-González (2014: 141). Initially, the scholar identifies two principal macro-categories of AVT research: conceptual research (which is more interested in exploring ideas and conceptualising them, rather than in scientific enquires of data) and empirical research, on which more attention is paid. Empirical research methods are data-driven, meaning that they require data from which hypotheses and generalisations can be formulated. Since data can be of various kinds, such empirical methods are further divided into different sub-categories. On the one hand, there are documentary research methods, which consist of seeking information within data produced for other purposes and on other occasions. Examples of documentary research are archival research[76], netnographic research[77], and corpus-based research[78]. In addition

[76] Archival research methods consist of accessing disparate documentary sources (e.g. original and translated copies of films or film dossiers) physically stored in archives or similar sites. There can be several reasons behind the study of these documents. For instance, they may be used to identify the impact of political censorship through subtitles or dubs (Pérez-González 2014: 161). Zanotti (2018) also mentions these methodologies for the study of past audiences' reception, using sources such as press reviews, box-office figures, fan magazines, or even letters or online comments.

[77] As the advent of the Internet meant that face-to-face interactions were often replaced by computer-mediated interactions in digital spaces such as forums, social networks groups or specific apps (like Letterboxd, in the context of film buffs), archival methods underwent a change. Indeed, the term 'netnography' refers to the immersion in virtual communities and the collection of (either pre-existing or elicited) data. Several studies have been conducted, for instance, with the objective of examining fansubbers' environment and practices (e.g. Li 2019).

[78] Corpora are among scholars' preferred tools for the qualitative and quantitative analysis of spoken or written screen dialogue., as they aim to provide authentic data and to go beyond the study of single case studies (see, for example, Pavesi 2019).

to this, author identifies two distinct research methods, observational and interactionist, between which a line separating perception and reception can be traced. Indeed,

> "observational research methods, illustrated by eye-tracking studies, address the perceptual and cognitive aspects of audiovisual translation, with a view to inform professional practices and achieve a better understanding of how viewers process information in screen-based environments. Interactionist research methods, involving the use of questionnaires or interviews, bring to the fore important considerations pertaining to the reception of audiovisual translations" (Pérez-González 2014: 141).

Similarly, Kruger and Doherty (2018) address the issue of audience reception of AVT products distinguishing between online and offline measures. In this perspective, offline measures are traditionally used in audience reception to obtain qualitative information (using, for example, questionnaires and interviews), while online measures are "typically more objective and physiological, including eye tracking, electroencephalography, galvanic skin response and heart rate" (2018: 97). From these two positions, the difference between the two terms mentioned at the beginning of this section becomes clearer. Similar yet different in many respects, both refer to the viewer experience but while perception could be summarised as the 'sensory experience', reception is rather the effect that a specific audiovisual text has on the audience. According to Gambier (2018), "studying reception [thus] means to investigate the way(s) in which AV products/performances are processed, consumed, absorbed, accepted, appreciated, interpreted, understood and remembered by the viewers, under specific contextual/socio-cultural conditions" (2018: 56). Referring to the AVT mode of subtitling, Gambier identifies three different types of reception ("the 3 Rs"): response, reaction and repercussion. Response concerns the perceptual decoding of subtitles from a psychological perspective (e.g. how we distribute our attention and why); reaction mainly aims at analysing viewers' processing effort from a cognitive perspective (the greater the effort, the lower the understanding – and thus, the quality – of the translation); repercussion, on the other hand, deals with preconceptions and consequences, i.e. it aims to reveal preferences, attitudes, habits regarding a given AVT mode, in a sociocultural dimension as well (2018: 57). It is cautiously suggested here that these 3 Rs, perhaps with some modifications, can also be considered valid in the field of dubbing.

3.1.2. Reception theory

Studies on reception belong to "a shifting area, with different paradigms in the humanities and social sciences for ways of understanding how and why people respond, or participate in the media" (Hill 2018: 3). The theoretical framework on which most of such studies are based is the reception

theory, which developed as a reaction to the traditional conceiving of the audience as a passive entity and aims at discovering the audience's active role in the construction of the meaning of a text. The theory is not strictly related to a single disciplinary field. On the contrary, it can be studied through a multitude of lenses, including sociology, psychology, pyscho-neurology, economics and linguistics, to name but a few. One of the first scholars to talk about (audience) reception theory was the cultural theorist Stuart Hall, who, in his 1973 paper titled *Encoding and Decoding in the Television Discourse*, proposed a new view on how (audiovisual) texts are produced and, later, interpreted by viewers. Following this approach, viewers are not passive recipients but active decoders that understand texts relying on a number of different factors related to their social and cultural background. This consequently means that texts are inevitably destined to multiple meanings on the basis of the audiences' interpretations. In effect, Hall states that "reception of the television message is thus itself a 'moment' of the production process" and that "production and reception of the television message are, not […] identical, but they are related: they are differentiated moments within the totality formed by the communicative process as a whole" (1973: 3). As evidenced by these quotations, the theory originated in conjunction with television studies. Indeed, the history of reception studies in film is relatively recent, given that until the mid-1990s, films were mostly examined through text-oriented approaches (i.e. with a focus on the authors, the content, or the form, as for literary texts) (Biltereyst and Meers 2018: 22). In parallel, issues of reception were investigated for what concerned television, conceived on one side as a novel technology, and as a social medium on the other, "hence implying audiences consuming and being influenced by television's outputs in a socially embedded environment" (2018: 22).

In fact, during the first half of the 20th century, there were experimental audience-based studies that can be seen as forerunners of modern reception studies. In the early days of cinema history, viewers were interviewed to study the possible psychological or societal dangers associated with the new medium (2018: 23). Also, for instance, studies were conducted by Hollywood film companies to determine the translation mode preferred by the audience at the time of the transition to sound (Zanotti 2018: 136). Apart from these experiments, however, for a long time, film studies as an academic discipline remained alien to the concept of the active viewer. It was only at the end of the 20th century, as already mentioned, that the audience was incorporated, so that "it was no longer solely the text that builds the identity of viewers" (Biltereyst and Meers 2018: 25), but also the viewer who builds the meanings of the text. In this context, some empirical reception studies also examined viewers' preferences, or how the social context of the audience can determine the success of films. Austin's case study (2002) is an example of this type of research: analysing the distribution of three adult-oriented films – i.e. *Basic Instinct* (Verhoeven 1992), *Bram Stoker's Dracula* (Coppola 1992) and *Natural Born Killers* (Stone 1994) – the author explored how the film industry can manage to anticipate patterns of reception. Referring to readers (although the concepts can

also be applied to viewers), Gambier explains how they "are not naïve: they bring 'scripts', 'schermata', previous knowledge, ideology, prejudices, experience etc. with them" (Gambier 2018: 46). This means that the audience select *a priori* what it wants to be exposed to on the basis of all these parameters (and, thus, also of what it has observed: film title and poster, trailers, reviews, and so on). Successively, there is the decoding and interpreting step, which is strictly individual. It is clear, therefore, that text meanings are not fixed entities established by creators, but rather something emerging from the constant exchange between creators, text and audience. Translation constitutes, then, an additional level of complexity, a delicate process which has the potential to radically change the overall meaning of a work. Hence, let us look now at the relation between reception and AVT, an aspect of primary importance for the present study.

3.1.3. Reception and audiovisual translation

In order to explain the importance of audiences for AVT studies, it is necessary to provide an account of how reception was first incorporated into the broader field of translation studies. In effect, a few years before Hall's publication of his abovementioned text, which was destined to change the way film studies was conceived, Eugene Nida published the ground-breaking *The Theory and Practice of Translation* (1969), a text of fundamental importance for the development of translation studies. In this seminal work – related to literary and Biblical translation in particular – Nida formulated a distinction between a pair of expressions: 'formal equivalence' and 'dynamic equivalence'. In formal-equivalence approaches, translators aim at rendering the target text while remaining as close as possible to the source text, in both content and form. On the other side, the dynamic-equivalence approach does not aim at obtaining equivalence in the text itself, but rather in the effect that the message has on the recipient. From an operational point of view, this means that translators have to find solutions sounding as natural as possible in the target language, so that, even if with important variations from the source text, "the response of the receptor is essentially like that of the original receptors" (Nida and Taber 1969: 202). Even though Nida's main focus was on the quality of translation, the text had a significant impact on the introduction of reader-oriented approaches in translation, and consequently, receptor-based studies. Indeed, Gambier (2018) identifies this contribution as the first step towards a systematisation of the active role of the recipient (reader, in the case of Nida), who becomes, in a sense, also a "reviewer" of the work carried out by translators (2018: 44).

As Chaume (2007) correctly suggests, "all interaction between a text and its receiver is directed towards a response from the latter" (2007: 71). Hence, given the growing importance of audiovisual texts as text types, audience reception has gradually turned into a major issue in the context of AVT, where the product of the translation process *per se* is no longer the sole object of

analysis. Talking about the multimodal nature of audiovisual texts (Section 1.1.1.), we have seen how analysis focused on language only cannot fully understand meaning, as there are several different ways of making meaning. In order to understand how meaning and other non-linguistic aspects are received (i.e. interpreted) by viewers, reception studies are crucial. When AVT is concerned, films are not the only text types to be studied, as even localised video games have been studied from a reception point of view (e.g. Mangiron 2018). Indeed, as many scholars point out, "there is no doubt that reception studies are essential to understanding the acceptability of AVT among audiences" (de los Reyes Lozano 2023: 10). Observational and interactionist research methods (or, alternatively, online and offline measures) seen above, are all used in the field of AVT reception studies, which is currently experiencing a period of great vitality due to the unprecedented growth in media production and consumption. However, as Di Giovanni (2022) explains,

> "from a translation studies perspective, the reception of translated audiovisual texts has only recently been made the object of systematic investigation, spurred by the ever-more frequent recourse to technologies for behavioural and psychological measures such as eye tracking, electroencephalography (EEG) or galvanic skin response, but also fuelled by the growing sophistication of long-standing tools like questionnaires, today administered in a host of different ways" (2022: 400).

In light of what has been said so far, it is possible to identify several studies in literature examining the reception (or the perception) of various AVT modes. In the field of subtitling, for instance, reception studies investigating cognitive load or comprehension have traditionally been used to evaluate the effectiveness of subtitles. Romero-Fresco (2020) provides a comprehensive overview of some of the most important reception studies carried out in the field of subtitling for the deaf and hard of hearing. This body of research – the author explains – was particularly important in providing evidence on the preference of subtitles over other accessibility modes (such as the sign language interpreting), as well as in establishing important parameters and guidelines regarding speed or format of subtitles. Moreover, today, new tools (mainly represented by the eye tracking technology) enable researchers to explore aspects such as the audience immersion. In fact, reception studies can be designed using the triangulation technique, which is valuable for cross-validation and "realised by combining two or more research methods for data collection […] in a serial or simultaneous manner in the study" (Kruger and Doherty 2018: 92). An example is Orrego-Carmona's (2016) reception study which, through both qualitative and quantitative research methods (questionnaires, interviews and eye tracking), aimed to find out whether Spanish viewers noticed differences between professional and non-professional subtitling in terms of overall understanding of the film plot (showing, eventually, good quality levels of amateur subtitles which

did not affect audience reception negatively). Another interesting example is Desilla's (2014) reception study, which investigated audience understanding of culture-bound references present in British films subtitled in Greek. The combination of qualitative and quantitative methods was also used in this case, with five-point scale questionnaires as well as with open-ended questions. Another recent contribution on the reception of subtitles is the study by Guerberof-Arena et al. (2024). In this case, researchers used questionnaires to examine the reception of a Mexican telenovela translated into English using three different modalities (human-translated, professionally post-edited and machine-translated). Results showed that machine-translated subtitles obstruct viewers' understanding and overall enjoyment.

In the context of audio description, for instance, Romero-Fresco and Fryer (2013) analysed the technique of audio introduction, which consists of a preliminary description of approximately ten minutes that incorporates information about the film (e.g. cast composition, production details, short synopsis, characters and setting overview, as well as information on the visual style, a much-debated topic also discussed in Section 1.2.2.4.). Reception was investigated in blind or visually impaired participants through questionnaires and a positive response to audio introduction was evidenced at the end of the study.

Finally, attention must be paid to the field of dubbing. As illustrated by Di Giovanni (2018), although dubbing is one of the oldest and most used AVT modes, it is still under-researched in terms of its reception by audiences (2018: 159). This is mainly due to two reasons. On one hand, research on dubbing has, hitherto, mainly been focused on linguistic and translational issues, or dubbing-specific constraints, adopting descriptive and contrastive approaches (2018: 160). On the other, studies looking at audiences' experience of dubbed audiovisual texts have, so far, seemed to be interested in perception (i.e. sensory processes) rather than in the broader interpretation or comprehension, which "pertain to the realm of reception" (2018: 161). As we have seen in the previous chapter, dubbing is experiencing a period of significant virality, given the globalised audiovisual market and the growing demand of localised content, the rapid advancements in technology and digital broadcasting, and the bottom-up AVT practices (i.e. amateur dubbing and cyberdubbing in general). As Pavesi clarifies, "viewers are not only becoming increasingly involved in the production of dubbing translations, but their role as consumers is now also receiving more attention from scholars" (Pavesi 2020: 160). The present study follows this trend, as it aims at analysing the reception of new forms of cyberdubbing. In general, the amount of reception studies on dubbing is recently increasing. Pavesi and Zamora (2022), for example, conducted a questionnaire-based reception study with the objective of studying swearing in film dubbing. In particular, the work investigated the degree of tolerance and acceptance of Italian and Spanish audience towards swear words and expressions in domestic and dubbed films. In this case, through an audiovisual survey containing series of clips, the comparison confirmed that Spanish viewers

are more tolerant, and that Spanish films (both original and dubbed) generally contain more swear words than Italian productions. Another example of a reception study involving dubbing was carried out by Perego et al. (2018), through a triangulation of questionnaires and other tasks. The research compared dubbing and subtitling, in order to assess which AVT mode affect cognitive effort the most when watching films considered complex. Results showed that, although AVT mode choice has little impact on the understanding of moderately complex films, it can affect the cognitive and evaluative processing in the case of more complex films – questioning subtitling in favour of dubbing. To conclude this brief overview, on the other hand, we could mention González Ruiz and Cruz García's (2021) reception study, in which the alleged "culture-neutralizing effect of dubbing" (2021: 223) was investigated. It is often held that dubbing diminishes source texts' foreignness, because, "unlike subtitling, the process of dubbing does not give the audience the opportunity to fully perceive the cultural gap between what they hear and see, and their own reality" (2021: 219). To investigate this aspect, two separate groups of participants were shown respectively a subtitled and dubbed version of the same film in parallel. After, they were asked to ascertain the national origin of the film they had just watched through a questionnaire. This was done to observe if there was any variation in the answers according to the AVT method used. However, results showed that there are more preponderant parameters for the identification of a film's origin, namely "its production values (i.e. Hollywood stars v. little-known European actors; blockbusters v. low-budgeted works […])" (2021: 229).

It is clear that audience reception, even just with regard to dubbing, is a broad field open to many types of investigation, and this is the reason why it was chosen for the present study.

3.2. Materials and methods

3.2.1. Video selection

Before discussing the empirical examination conducted here, it is appropriate to begin with an in-depth description of the materials used for the study in question.

As previously stated, the practice of dubbing from other languages into English is a relatively rare phenomenon. In section 1.3.2. we mentioned examples of English-language dubbing in the past, which however remained scarce in number until recently. Indeed, it has been previously confirmed that we are witnessing a rebirth of English-language dubbing in contemporary times. This can be observed in both professional and commercial settings (thanks to online streaming platforms), and in the realm of amateur online content creation, where the utilisation of new technological tools for translation is also becoming increasingly prevalent.

To answer to the RQ, it was necessary to collect excerpts of non-English-language films that had been dubbed into English and then submit them to native English speakers to evaluate a series of aspects. In this case, it was opted for original Italian productions dubbed into English.

In order to prevent the audience reception from being distorted by the viewing of similar clips with a high degree of uniformity in the dubbing (e.g. only clips dubbed with AI by amateurs), the material chosen is of varied nature, i.e. each clip chosen was dubbed at different times and in different context. In particular, four video clips were selected for the reception study. Two of these clips feature scenes from Italian films that have been dubbed into English by professionals (and have been officially released in the market), while the other two clips feature scenes from Italian films that have been dubbed into English by amateurs and through the use of AI-based tools.

With regards to the films dubbed by professionals, the choice was made by consulting the 'Italian films' subsection of Dubbing Wiki[79], a fan-based online resource that provides a comprehensive list of all Italian films dubbed into English from past to present. As it can be observed consulting the page[80], the list is not long, as it includes only 103 films. Although the page is maintained by fans and therefore there may be inaccuracies, the low number testifies the fact that the English-language dubbing of Italian films is not a historically consistent phenomenon. Indeed, if 103 may seem a reasonable number, it is considerably low if we look at the number of original Italian films that are produced each year. To give an example, in the year 2021 alone, after the Covid 19 pandemic, 313 film were produced in Italy[81].

Two further details emerge from looking at this list. First, 33 out of these 103 films are animated films produced in Italy. Second, apart from animated films, 29 out of these 103 films are produced and/or distributed by Netflix. For what concerns the remaining films, in most cases, they are films from the 1960s or 1970s, dubbed in the period of weakness of the US film industry that was mentioned in Section 1.3.2. On the one hand, these details confirm the limited diffusion of the English-language dubbing up to the present day, if we do not consider the animation genre, which is (as it has been already mentioned in Section 1.3.1. as an index of the vitality of dubbing) an exception. On the other, the significant acceleration provided by Netflix's marketing strategies is highlighted. Indeed, Netflix appears to be the only company involved in this 'revolution', while there is almost an absence of other OTT players distributing and dubbing Italian films into English.

[79] Dubbing Wiki is part of the larger website Fandom. Fandom is a "network of fan-curated wikis – a one-stop shop for fans of all types to explore and discuss their favorite stories, characters, and lore" (see https://www.fandom.com/), a sort of forum containing fan communities of films, TV series, video games and other cultural phenomena. On its part, Dubbing Wiki is a subsection created with the purpos of gathering "as much information about dubbing of productions into the English language" (see https://dubbing.fandom.com/wiki/Dubbing_Wikia).

[80] See https://dubbing.fandom.com/wiki/Category:Italian_Films.

[81] See https://cinema.cultura.gov.it/notizie/pubblicato-il-report-tutti-i-numeri-del-cinema-italiano-anno-2021/#:~:text=La%20Direzione%20generale%20Cinema%20e,%2Dpandemia%20(325%20nel%202019) .

The abovementioned list was thus employed to identify films that had been dubbed by human professionals. To provide some variety, one film dubbed in the past and one film distributed internationally by Netflix were traced. In the first case, an excerpt from *La vita è bella* (Benigni 1997) was selected. The film, about a Jewish Italian man trying to protect his son from the horrors of the Nazi concentration camp, combines drama and comedy creating an innovative Holocaust narrative.

This choice was made with the specific intention of identifying an example of English-language dubbing that was neither from the 1960s or 1970s, nor contemporary, as both ages are marked by very specific commercial traits. This means that the dubbing of this film constitutes, in some way, an exception, i.e. it was created during a period when the practice of English-language dubbing was not as prevalent. The reasons behind the decision to dub the film into English was, however, of a commercial nature in this case too. Following its release in Italy in 1997, it was presented at the Cannes Film Festival in 1998 (winning the jury's Grand Prix), and at the 71st Academy Awards, where it was bestowed with (among other prizes) the Oscar for Best Foreign Language Film)[82]. Other than a critical success, the film was an international commercial hit as well, as it represents the highest-grossing Italian film in history (generating a worldwide box office revenue of over $230,000,000)[83], and the second highest-grossing non-English-language film in the US (after the Taiwanese *Crouching Tiger, Hidden Dragon*[84])[85]. After the success of the English subtitled version in the English-language markets, the Miramax distribution company decided to re-release an English dubbed version of the film in August 1999, in what has been defined as the "Miramaximizing" effect (McDonald 2009: 363). However, many reviews highlighted a negative reception from film critics. For example, in a *Variety*'s article, Koelher (1999) declared: "the cause for dubbing foreign-language films into English for the U.S. theatrical market will receive little momentum from Miramax's Anglicizing of Roberto Benigni's Oscar-winning 'Life Is Beautiful'". The main issue ruining the film's enjoyability, in that case, seemed to be the lip syncing, which the American audience and critics were not accustomed to: "the dubbed 'Life' will thrive better on home screens than on big ones, where unavoidable dubbing mismatches are much tougher on the eyes and ears" (1999). Similarly, Kaufman's (1999) editorial on *IndieWire* levelled criticism on the choice of dubbing "a motion picture that did just fine in the first place". The criticism was not only focused on the film company's intention to reach a wider audience in order to monetise, but also on the dubbing itself. This is what the author says about dubbing actor Jonathan Nichols performance in dubbing Roberto Benigni with a marked Italian accent, and about dubbing in general:

[82] See https://www.mymovies.it/film/1997/lavitaebella/premi/ .
[83] See https://www.boxofficemojo.com/title/tt0118799/?ref_=bo_se_r_1 .
[84] Original title: *臥虎藏龙*; pinyin: *Wòhǔ Cánglóng* (Lee 2000).
[85] See https://www.boxofficemojo.com/genre/sg4208980225/ .

> "Nichols's Italian accent isn't always right and his inflections are not nearly as zany as Benigni's, but after about an hour, I began to get used to it. After all, most of Benigni's humor is physical anyway. But […] no matter how good the dubbing, it's still dubbing and it is still a lesser experience. […] Thankfully, we are not used to it, nor should we get used to it. The 'stigma against dubbed films' is a natural one. It's like saying you have a stigma against Cheese Whiz. It's artificial and doesn't taste as good. If Weinsten [i.e. chairman of Miramax] cares about foreign movies so much, how come he's making the Cheese Whiz instead of fostering the Brie?" (Kaufman 1999).

If these two instances could be seen as representative of "the broad critical rejection of the dubbed version" (McDonald 2009: 372), audience's rejection of dubbing as an AVT method is still interesting to investigate in contemporary times, and will be addressed in detail later during the reception study.

In addition to these sociocultural factors, this film was selected for the translation study because another scene from the same film was dubbed using AI tools by the Instagram project 'italiancomedydub'. Comparing the two dubbed versions of the film was, thus, an opportunity to gather authentic information about audience's ability to distinguish between synthesised speech and speech produced by dubbing actors.

The second film dubbed by human professionals selected is *L'incredibile storia dell'isola delle rose* by Sydney Sibilia. A scene from this film was chosen as it can be representative of the new Netflix-mediated English dubbings made available in recent times. Indeed, the film was produced by the production company Groenlandia (co-founded by Sibilia himself) and distributed by Netflix on the platform from December 2020. The English dubbing was commissioned to the Los Angeles-based dubbing studio Post Haste Digital (later acquired by the bigger company Deluxe)[86]. The film tells the true story of the engineer Giorgio Rosa (played by Elio Germano) who, during the protests of 1968, built a platform (called 'Isola delle Rose') in international waters and declared it a nation. In this case, the film is not particularly interesting from a commercial or critical success perspective, but because it is coherent with the material for the reception study, as it also belongs to the comedy genre.

The other two clips selected for the study are, indeed, also instances of comedy cinema. Produced by the joint work of the admins of 'italiancomedydub' and the followers involved in the creative project, who are "at the same time, consumers and producers of translations" (Orrego-Carmona 2018: 324), the two scenes are exemplar of the AI-powered English-language cyberdubbing under investigation here. Although the process is not entirely automatic (as specified

[86] See https://dubbing.fandom.com/wiki/Rose_Island .

in Section 2.3.), these videos will still be referred as 'AI-dubbed videos' in the thesis and in the interview. The reasons is because, already mentioned in Section 2.2., 'AI' is an umbrella term covering a wide range of machine-learning-based task. Wu et al. (2023) explain that there are usually two scenarios for video dubbing: one is the direct translation of the speech of a video from one language to another (i.e. what we could define as a fully automated dubbing); the second scenario consists in generating speech according to a text transcript (2023: 13772). In our case, the label 'automatic' will be used referring to the second scenario, i.e. the speech synthesis stage.

With regard to the clips, the selection was made by consulting the 'italiancomedydub' Instagram page. As already mentioned, one scene is from *La vita è bella*, while the other is an excerpt from the popular Italian film *Tre uomini e una gamba* (Baglio et al.) who came out in the same year (1997) as Benigni's dramedy[87]. The film is a comedy that revolves around three friends (played by the trio of comedians Aldo, Giovanni and Giacomo) who embark on a trip across Italy to deliver a sculpture (a wooden leg) to their boss.

The page mainly publishes its content in the format of short vertical clips, which are more suitable for mobile devices screens and social networks interfaces, as well as more convenient to access, since the timing online is rapid and accelerated. In addition to this, the videos show the page's logo, the caption 'doppiato con IA by ItalianComedyDub', and subtitles in Italian (since, as already said in Section 2.3., the followers are mostly Italian native-speakers who enjoy watching their favourite films dubbed into another language). However, the page's admins were contacted in private, and the AI-dubbed videos were made available in the original film format, and without caption and subtitles, in order to provide interviewees with video material that did not seem to be specifically designed for the purpose of entertainment on social networks.

As will be seen in the following sections the four scenes could be classified as comic[88]. This was done for two reasons. First, the context of comedy is traditionally prone to a number of typical translational problems associated with AVT in general. Examples of such problems are provided by Delabastita (1990: 102), who mentions, among the others, the rendering of wordplays and humorous language, the rendering implicit cultural references, and the rendering of particular language varieties or personal idiosyncrasies like speech defects (which are often used as a tool to obtain comic effects). It seemed interesting to investigate the treatment of these issues by professionals and amateurs, as well as the reception of these elements by an audience which is not familiar with the practice of dubbing.

[87] 'Dramedy' is another term used to designate works that combines elements of comedy and drama at the same time.
[88] Even though in the case of *La vita è bella* the context is dramatic, the scenes have a humorous intent.

In the second place, the choice was due to the intention to ensure uniformity through a lateral evaluation, i.e. by comparing similar[89] but not identical scenes. Indeed, a direct comparison between the same scene dubbed by humans and dubbed using AI, could have led to a predictable result, i.e. interviewees, presented with the same clip repeated twice, would have perhaps been more inclined to perceive one as qualitatively different from the other. Instead, as will be summarised later in the section dedicated to the description of the procedure of the research, the selection of four different scenes permitted, in some cases, for the interviewees to remain unaware of the artificial nature of part of the material.

The clips, which have variable duration (average duration: three minutes), will now be briefly described in the following sections, in order to facilitate the comprehension of the answers to the interview in the next chapter, when references to the content of the scenes will be made.

3.2.1.1. Clip 1

The first clip presented to the respondents during the reception study is an excerpt from the English (human-) dubbed version of *La vita è bella*. It consists of a particularly tragicomic scene that could be divided into two parts. In the first part, Guido (the protagonist, played by Roberto Benigni) and his little son Giosuè arrive at the Nazi concentration camp. The scene is set inside a barrack, which is crowded with many other deported Jews. Giosuè asks his father to see his mum and complains of being hungry. At this point, Guido lies to his child to comfort him, saying that food will be served soon. It is the beginning of a game of imagination that the father will use to protect his son from the horrors of the camp. When a Nazi officer arrives, asking for someone to translate the camp's rules from German to Italian, Guido volunteers even though he does not speak the language. This could be seen as the beginning of the second part of the scene. Here, Guido starts using his humour and imagination by translating the officer's speech in its own way, i.e. explaining the rules of an invented game for his son instead of the actual camp's rules.

3.2.1.2. Clip 2

The second clip is a scene from the comedy *Tre uomini e una gamba*, dubbed using text-to-speech software by the 'italiancomedydub' project. The three male protagonists (middle-aged men) are sitting at an open-air restaurant table, having a conversation. In this case too, the scene in question could be seen as divided in two parts. In the first part, Aldo and Giovanni mock their friend Giacomo, who appears to have developed romantic feelings for a woman, despite his imminent wedding. When the woman in question joins the table, a rapid dialogue informs the viewers that

[89] The adjective 'similar' is used in this case to refer to the filmic genre, which is comedy in all cases. For what concerns *La vita è bella*, even if dramatic elements are present, the scenes selected for the reception study present jokes and humorous dialogue.

the woman, who is visibly moved by the topic, had been recently dumped by her boyfriend. Other dialogues (about the wooden leg and about the men's profession) alternate, along with quick shot/reverse shots. The second part of the clip differs from the first in terms of composition and overall scene direction. Shots are slower and less close-ups are present, as the camera moves around the four characters seated at the table. Additionally, a gentle music begins to play, as the woman elucidates a Plato's philosophical concept to the three men.

3.2.1.3. Clip 3

The third clip shown during the interview is from *L'incredibile storia dell'isola delle rose*, and constitutes one of the two examples of English dubbing made by human professionals. The scene is set inside an unusual car. Inside the vehicle are Giorgio (the protagonist, played by Elio Germano) and Gabriella (Matilde De Angelis). The conversation between the two is playful and provides some details: it is revealed that Giorgio built the car for his engineering exam and that the two characters had a relationship in the past. As the scene draws to close, a police car pulls up, as the Giorgio's 'invention' does not have a licence plate. Following Giorgio's explanation, an officer instructs him to exit the vehicle.

3.2.1.4. Clip 4

Clip 4 is the second instance of AI-powered dubbing. It consists of another scene from *La vita è bella*. This time, however, the revoicing was not carried out in professional studios, but in amateur settings and with AI tools, i.e. by the 'italiancomedydub' project. The clip is short and consists of a dialogue between Guido and Giosuè (father and son). The two characters, in this case, are not in the concentration camp, but are casually walking through the streets of an Italian city. They are well dressed like the other characters walking down the street, except for the troops of marching soldiers, who signal the particular historical context. When the child approaches a café window asking his father for a cake, the camera shows a sign that states "Vietato l'ingresso agli ebrei e ai cani"[90]. In response to his son's curiosity, Guido says that each shop has its own unique rules. Walking away, the father consoles his son by telling him that, from the next day, their bookshop will adopt a similar policy, prohibiting the entry of spiders and Visigoths.

3.2.2. Methodology

Within the field of AVT, many aspects can be studied through a variety of different lenses. For what concerns reception, it is important for researchers to choose the most appropriate approach

[90] English: "No Jews or dogs allowed".

to gather useful data and build or test some hypothesis. Indeed, selecting the right methods is perhaps the most important step in empirical research. Reception in AVT has already been described in Section 3.1.1. – with regards to Gambier's (2018) "3 Rs" – as a broad field encompassing all aspects of the audience's film-watching experience (e.g. how AVT modes are decoded and interpreted by the audience from a sensorial, cognitive, or sociocultural standpoint). To address this macro-topic in a narrower manner, it should be noted that audience reception will be analysed here through what Pérez-González (2014) defined as "interactionist research", and using the methodology proposed by Kruger and Doherty (2018) which involves the use of the so-called "offline measures", i.e. questionnaires and interviews. In particular, the empirical investigation (the procedure of which will be described below) will entail the viewing of the abovementioned clips, followed by an interview (see Appendix B), i.e. a series of open-ended questions on various themes designed to elicit the audience's direct response to the clips. As it will be outlined later, the last questions of the interview aims to investigate general opinions and attitudes towards the research topics. For this reason, it could be stated that the initial part of the interview concerns the second "R" (reaction), while the final part of the interview concerns the third "R" (repercussion). Indeed, as Tuominen (2018) explains with regards to the terms proposed by Gambier:

> "While response is clearly connected to an individual viewer's ability to follow a translation, both reaction and repercussion are more contextually oriented: reaction covers the immediate context of viewing and of individual interpretations, and repercussion looks at the broader context of audiovisual translations as a factor in their viewers' lives and in society overall" (2018: 70).

Tuominen also underlines how these aspects have traditionally been more investigated through questionnaires, as opposed to interviews (2018: 71). This is because questionnaires allow researchers to manage quantitative data (obtained, for instance, through Likert scales[91]), which are easier to analyse and typically more objective. If particularly structured, questionnaires can be effective methods for confirming pre-existing theories. Indeed, by presenting respondents with closed questions and pre-elaborated answers, researchers can direct the study and verify the accuracy of their initial hypotheses. On the other hand, interviews could be regarded as a form of

[91] Likert scales (named after their inventor, the psychologist Rensis Likert) are also commonly known as rating scales. These scales represent the most widely used method for measuring the opinions or attitudes of participants to a survey or questionnaire. The scales were initially devised within the field of social psychology but are now employed in a multitude of research domains.

qualitative investigation[92]. This is true not only because interviews present open-ended questions that allow the participant to express – for instance – evaluations or judgements in a more spontaneous way, but also because their theoretical framework enables researchers to gather new information that was previously unexplored. It is thus necessary to determine whether the research is regarded as a means of corroborating a theory, or as a tool for obtaining data to construct new hypothesis. In this case, interview was chosen as the research method for two main reasons. First, research topics are positioned in a particularly novel field, since the spread of AI tools, the consolidation of cyberdubbing cultures and the establishment of English-language dubbing are all contemporary phenomena linked to technological development and online domains. Therefore, the interview was designed with the objective to elicit data on (partially) unexplored themes. Second, since the topic of AI's application in AVT is a broad field, this study could be regarded as a preliminary exploration. Indeed, as Chaume (2018) suggests:

> "qualitative research in AVT is used to obtain a deeper understanding of the underlying reasons, opinions, and motivations of all the agents involved in the process of translation, be it distributors, translators, or even the audience. It provides insights into any given issue discussed, and entices the development of ideas or hypotheses for potential quantitative research" (2018: 51).

Different research methods are not mutually exclusive: many studies triangulate data analysis methods. In this case, future research could build upon these findings to confirm trends that will be identified here though larger-scale quantitative analyses.

3.2.2.1. Participants in the reception study: demographics and viewing habits

As already mentioned, the present study was conducted on a small scale and aimed at investigating English native speakers' reception of a particular kind of English dubbed content. The participants recruiting process started within the Linguistics department of the University of Pavia. Three American English native speakers were identified, contacted privately, and asked to participate in the interview. Subsequently, a 'snowball sampling' approach[93] was used in order to recruit other subjects for the research. At the end of the recruiting process, ten subjects were selected to join to the qualitative interview. In all cases, respondents took part on a voluntary basis.

[92] Interviews can be unstructured or semi-structured. When an interview is semi-structured, the respondents are guided by the researchers, e.g. with fixed pattern of questions. Nevertheless, such interviews are still to be considered as less restrictive than questionnaires, which allow for more precise but less spontaneous answers.

[93] Snowball sampling (also referred to as chain sampling) is a sampling technique where existing participants to a study recruit other subjects to be tested on the basis of their personal acquaintances.

Despite the limitations implied by a small sample size, it was here posited that a group of ten participants could be sufficient, especially because of the depth of data that can be generated through open-ended questions. As already mentioned, indeed, long and detailed responses allow for more nuanced explorations of research themes and can constitute the basis for future research.

While the importance of sample width may be limited in certain circumstances, the composition of the sample remains a crucial factor in all research. In the case of large samples, for example, it is important to ensure diversity and representativeness, as quantitative analyses typically aim at obtaining reliable statistics to formulate norms and uncover patterns. On the other hand, in the case of small samples, diversity could be an issue: due to the limited number of subjects tested, no statistical measures can be applied. For this reason, it is better to capture perspectives and data linked with a specific population. As reminded by Tuominen (2018),

> "despite careful research designs, these [small-scale] studies are not broadly generalizable: the interpretations of Italian audiences tell little of the interpretations of German or Brazilian audiences, and reactions to Hollywood blockbusters might be different from reactions to independent arthouse films. Therefore, we cannot take these studies as indicative of reception overall. Rather, the studies can attempt to provide somewhat reliable information on the population which they explore" (2018: 79).

In this case, the objective was to investigate a highly circumscribed population that could reveal interesting aspects on the research themes. The first questions in the interview (see Appendix B) consist of a brief the sociolinguistic overview of the participants to the study. This is because, as Gambier (2018) states, "different variables related to viewers are to be taken into consideration if and when reception is to be studied: age, sex, education background, […] frequency and volume of AVT consumption, AVT habits (opinion and preference), command of foreign languages, […] etc" (2018: 53). Answers to these preliminary questions confirmed the original intention to collect answers from a small group of subjects that was however coherent inside and that could allow to gather data to be confirmed in the future by updated research.

Indeed, the sample is composed by English native speakers only (two British English native speakers and eight American English native speakers). The age of the respondents ranges from 23 to 33 years old, with an average of 26.9 years old. The education level of the group is generally high: six of the ten respondents have obtained a master's degree, while the remaining four have a bachelor's degree. Six of the ten respondents study as their main occupation, while three work and one is unemployed. All the respondents also speak another language. In eight of these cases, this language is Italian, although with different levels in terms of speaking, reading, writing and overall comprehension. Even though this aspect is based on self-evaluation, it is important because it could

affect the reception of the dubbed clips, since the source language is Italian in all cases. Command of foreign languages is also significant as it anticipates the questions on frequency and volume of audiovisual content consumption and on AVT habits. In particular, the interest towards films and TV series is widespread: nine of the ten respondents declare to engage with these forms of entertainment at least once a week, while four consume such audiovisual products on a daily basis. For what concerns the national origin and the original language of the productions, the answers showed heterogeneity: four respondents declare to watch mostly English-language content, while the rest of the interviewees does not have a clear preference and often enjoys foreign productions[94]. With regards to the favourite AVT method, the answers to the preliminary questions show a definite preference for subtitling: all respondents say they use subtitles as a support when watching a foreign film or TV series. This aspect, and the related lack of tolerance towards dubbing will be investigated in greater detail in the next chapter.

The aforementioned characteristics (age, education, language competencies, viewing habits and preferences) provide insight into the composition of the sample used for the reception study. Indeed, as Table 3 below summarises, the sample could be visualised as young, educated, engaged with the consumption of audiovisual content, and tending to be unfamiliar with dubbing as an AVT mode. The importance of this parameters in the interpretation of the answer is not to be underestimated. Apart from individual judgments on translation modes *per se*, those who are young and educated are also likely to have good levels of computer literacy. All these aspects can play, thus, a major role in shaping the reception of dubbed content and of AI-generated products.

Table 3: Demographics and viewing habits

	Age	Education level	Occupation	Mother tongue (L1) and foreign languages (L2) spoken	Frequence of audiovisual content (film and TV series only) consumption	Favourite productions' origin (domestic vs. foreign)	Favourite AVT method
R1	27	Master's degree	Student and part-time designer	L1: American English L2: Italian, Spanish	At least once a week	Both domestic and foreign	Subtitling
R2	29	Master's degree	Student and English language teacher	L1: American English L2: Italian, Russian	Daily	Mostly domestic	Subtitling
R3	27	Master's degree	Student	L1: American English L2: Italian	Almost never	Mostly domestic	Subtitling
R4	25	Master's degree	Student	L1: American English	At least once a week	Both domestic and foreign	Subtitling

[94] In this sense, four respondents stated that they enjoy English and other language productions equally, while two others stated that they started watching content in another language (Italian) after moving to Italy.

				L2: Italian, German			
R5	26	Master's degree	Student and English language teacher	L1: American English L2: Italian	Daily	Mostly foreign	Subtitling
R6	26	Master's degree	Retail	L1: British English L2: Spanish, Italian	Daily	Mostly domestic	Subtitling
R7	23	Bachelor's degree	Designer	L1: British English	At least three times a week	Both domestic and foreign	Subtitling
R8	26	Master's degree	Student	L1: American English L2: Spanish	At least four times a week	Mostly domestic	Subtitling
R9	33	Bachelor's degree	Unemployed	L1: American English	Daily	Both domestic and foreign	Subtitling
R10	27	Master's degree	Accountant	L1: American English L2: Chinese, Spanish	At least once a week	Both domestic and foreign	Subtitling

3.2.2.2. Procedure of the reception study

Having outlined the composition of the sample that took part in the interview, it seems now appropriate to provide details about the procedural protocol adopted for the study, i.e. the steps that were systematically followed in all ten interviews.

After the recruiting process briefly described in the previous section, two additional volunteers were consulted to test the structure of the interview prior to the actual empirical investigation. This was primarily done in order to:

(i) verify the most appropriate order of the clips to be shown;

(ii) assess if the audience immediately recognised that some videos had been realised using AI-based tools;

(iii) evaluate whether the questions concerning AI and its impact should have been posed before or after the viewing of the clips;

(iv) ensure consistency and guarantee uniformity in the actual interviews' development.

Following this preliminary stage, interviews were conducted with the ten respondents using a multi-site approach, i.e. they were carried out both in person and in online settings. In particular, three participants were tested in person, while the rest was interviewed via a videoconferencing application. For this part, the app Zoom was used, as it allows users to share files such as videos or audio in real time. In all cases, the interview was individual, conducted in English, and took an average of 35 minute to complete.

Since ethical approval is essential for projects relying on questionnaires or interviews, each respondent was initially warned that the conversation was going to be recorded and later

transcribed for the purpose of a research. Hence, an informed consent form was provided to each participant (see Appendix C). For what concern the topic of the study, participants received a general introduction followed by the description of the procedure. No mention of the artificial nature of two of the four clips was made prior to the viewing session, in order to prevent participants to be influenced and biased in the viewing session.

As already mentioned, the initial phase of the interview involved some preliminary questions to ascertain the sociolinguistic background and the AVT preferences of the participant. The second phase consisted in the viewing of the four clips selected (following the numbering indicated in the previous sections). The clips were in all cases viewed on computer screens. This detail is of particular importance, since, as Tuominen (2018) states: "what is central [in these cases] is the creation of a viewing situation which allows the test participants to watch the translated programme in a normal, authentic way, and to construct whatever interpretations arise naturally from the viewing situation" (2018: 71). A key term in this context is that of 'ecological validity', which underlines the importance of the research setting. Indeed, "even though not generalizable, the research setting should credibly reproduce authentic conditions, so that it is possible to assume that the behaviours observed in the empirical setting are similar to behaviours in normal, everyday settings of the same type" (2018: 82). This study attempts to respect this concept, as cybverdubs are typically watched at home, through computer (or even smaller mobile phones) screens, and not in 'formal' contexts such as movie theatres, where other considerations could emerge, due to the significant differences in terms of audio and video quality (e.g. higher and more immersive volume, bigger screen).

After each clip, the respondent was required to answer the five 'Reception questions' shown in Appendix B, concerning the naturalness of the dubbed speech, the synchronisation, the emotional tone, and the overall comprehension of the scene. Since four clips were shown, a total of 20 'Reception questions' were posed to each interviewee. Therefore, for the reception part alone, interviewing ten subjects resulted in a total of 200 answers, which provided a broad set of insights for the analysis in the next chapter.

The final stage of the interview consisted of questions related to artificial intelligence and its effects. After some questions aiming to discover the respondents' familiarity with the themes of the research, a brief task was carried out. Indeed, participants were informed that two of the four clips had been revoiced using a speech synthesis software and were asked to identify them. In this phase, interviewees were not allowed to view the clips again and had to rely on their memory. As Di Giovanni (2017) highlights,

> "another key issue in effects research is that of memory, or recall, i.e. media users ability to retain and reproduce information from the media content they have been exposed to, often

> within a short time from exposure […]. Indeed, memory can be taken as a valuable indicator of what the viewers have been mostly impressed by, and what they have found most relevant" (2017: 165).

The choice behind the identification task's functioning was thus guided by the intent of bringing out the most relevant observations.

All interviews were recorded and transcribed automatically with an in-built feature of the recorder of the phone (Google Pixel 7) used for the study. Subsequently, the audio files were played back, and the transcribed text was revised. Answers were then edited to remove impractical elements such as repetitions, hesitations or interjections.

After this stage, the transcribed answers (along with the corresponding questions) were imported into an Excel file (*interview.xlsx*) to obtain a usable file which was quick to access in the data analysis stage.

3.2.2.3. Data analysis

Quantitative research, using methods such as the questionnaire, can be employed to examine the quality of dubbed products in a variety of ways. Spiteri Miggiani (2024), for instance, proposes a quality assessment model designed for "in-studio dubs, voice-overs, and AI-generated outputs" (2024: 68). The model, named "Script, Speech and Sound Quality Assessment Model" (2024: 57), encompasses two analytic sets of parameters: on one side there is the "Script Rubric", which focuses on textual elements and aims at discovering errors in contrast to the original text; on the other, the "Speech-and-Sound Rubric" is viewer-oriented, i.e. it involves an "acceptability score system to evaluate the quality of speech and sound elements in the dubbed product" (2024: 68). Since in this case the research will be based on qualitative interviews, the utilisation of such scores was not an option[95]. It was still necessary, however, to establish a framework for data analysis.

As a matter of fact, in order for qualitative research to be valuable, it is important that studies are conducted in a methodical manner, i.e. following an approach that is recognised, established and credible. Indeed, although qualitative studies are generally considered to be less generalisable than their quantitative counterparts, following a rigorous qualitative research method is useful to treat data in an optimal manner, shedding lights on unanticipated insights.

In the context of the present study, the data analysis will be based on the theoretical framework of the 'thematic analysis'. To elucidate this concept and to explain its functioning, this brief section will draw on an article by Nowell et al. (2017). In particular, thematic analysis is described as a "a qualitative research method that can be widely used across a range of

[95] Nevertheless, some quality indicators defined in the study will be taken into account for the analysis of the answers to the interview.

epistemologies and research questions" and as "a method for identifying, analyzing, organizing, describing, and reporting themes found within a data set" (2017: 2). The authors define six phases that precede and then constitute the qualitative data analysis. In particular, the third phase consists in "searching for themes" (2017: 8). First, the concept of 'theme' is clarified as a recurrent abstract entity that is related to the research questions and that links parts of the data to the research questions. Once data is collected, themes must be identified. When addressing themes identification, an important pair of terms is 'inductive' or 'deductive' thematic analysis. Inductive thematic analysis is data-driven, i.e. "the themes identified are strongly linked to the data themselves and may bear little relation to the specific questions that were asked to the participants" (Nowell et al. 2017: 8). In contrast, deductive thematic analysis is quite the opposite: with a top-down approach, themes are traced within the questions raised by the researchers. In this case, a deductive thematic analysis will be carried out, as the themes identified are connected to the questions posed to the subjects in the interview.

Not every question constitutes a theme: the fourth phase outlined in the article ("Reviewing themes") shows how themes must be reconsidered in many cases during the data analysis. If a new interesting topic is found within an answer that is not covered by any of the pre-established themes, then the set of themes can be expanded. On the other hand, if a theme is deemed to be particularly irrelevant due to the absence of pertinent data, it can be deleted. Finally, if topics overlaps, themes can converge. Therefore, even if a deductive thematic approach is used, there is no one-to-one correspondence between questions and themes. In our case, for instance, four themes (corresponding to the salient topics of the interview) will be identified in separate sections within the next chapter. In addition to this, the discussion will also unfold in some subthemes, to provide a comprehensive view of the answers provided by the respondents. As will be shown, the identification of specific patterns in the responses will result in the creation of the so-called 'codes'. As elucidated by Saldaña (2013):

> "a code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. The data can consist of interview transcripts, participant observation field notes, journals, documents, drawings, artifacts, photographs, video, Internet sites, e-mail correspondence, literature, and so on" (2013: 3)

In our case, the interview passages employed for the purpose of thematic analysis will be reported and broken down into specific codes. This is, as explained by Saldaña, an essential aspect of the work of qualitative data analasys, since: "when we reflect on a passage of data to decipher

its core meaning, we are decoding […] [and] when we determine its appropriate code and label it, we are encoding[96]" (2013: 5).

[96] More generally, "coding is the transitional process between data collection and more extensive data analysis" (Saldaña 2013: 5).

CHAPTER 4
RESULTS AND DISCUSSION

4.1. Perceived naturalness

The first group of answers to be examined here concerns the theme of 'perceived naturalness' of the dubbed speech. For each of the four clips, indeed, the first question (Q1) of the interview aimed to investigate the degree of naturalness associated with the voices of the dubbing actors. Following the first question [11], respondents were left free to apply their own interpretations of the concept of naturalness, as no further parameters were provided.

> [11] Q1: Did the voices in this clip sound natural to you? If not, can you describe when or why the voices felt particularly unnatural?

As already mentioned in Section 2.2.1. when talking about synthesised speech, "the obvious drawback of working with naturalness is that, by casting the net too wide, we are left with a very vague concept; one that is especially difficult to pin down and dangerously prone to trigger impressionistic observations" (Romero-Fresco 2009: 51). This study by Romero-Fresco (2009) could be mentioned as an example of how generally naturalness is treated in the field of AVT studies, since the notion is "wide enough to account for the different features of dubbing language" (2009: 51) and requires boundaries to be studied objectively. In that specific case, the concept was narrowed down and compared to that of 'ideomacity', i.e. the audiovisual translators' ability to select and use nativelike expressions in the target text. In contrast, the concept of naturalness in the present interview was not further specified, to leave respondents room for interpretation. This has indeed led to the emergence of many different conceptualisations of naturalness.

For what concerns *Clip 1*, i.e. the first scene from *La vita è bella* dubbed to English by professionals, while three respondents[97] did not mention any element as particularly unnatural and were overall pleased by the dubbing, the rest concentrated on a number of aspects related to the speech and to the dubbing actor's performance. As mentioned before, in this context, such characteristic and recurring aspects will be defined as 'codes'.

Interestingly, some answers focused on the lack of synchronisation between the dubbed speech and the lip movements of the actors as a key indicator of the dubbing's overall lack of naturalness. This is testified, for instance, by R1, who affirms that "the dialogue didn't feel natural

[97] From this point on, respondents will be identified and indicated with the abbreviation 'R' and the corresponding number. For example, respondent number one will be indicated as 'R1', respondent number two will be 'R2', and so on.

because of the lip movements". Apart for the lip sync issue, which will be addressed in more detail later, the remaining answers contributed to the emergence of other codes.

4.1.1. Accent

A problematic aspect that could be traced within the answers could be 'encoded' as ACCENT. As already mentioned in Section 3.2.1., the dubbing actor Jonathan Nichols opted for a strong Italian accent in the English dubbed version of the film. Apparently, this choice was a significant factor that prevented the dubbed speech to be perceived as natural. As Table 4 shows, this was highlighted by four of the respondents. This is of particular interest, especially if we consider what said in Section 1.3.4., in relation to the possible strategies when the dubbed speech needs to reproduce the source text's language varieties in the target language. With regards to this issue, Hayes (2023) identified four main strategies that the streaming giant Netflix has adopted in recent years (yet the considerations seem to be valid for dubbing companies in general). The first strategy consists of standardising the original texts' accents or dialects using quasi-artificial varieties in the target language. This aspect was abundantly described in Section 1.3.4., as it can be considered as a recurrent feature of the language of dubbing and concerns the so-called 'prefabricated orality', which Baños and Chaume (2009) defines as "an orality which may seem spontaneous and natural, but which is actually planned" (2009: 1). Indeed, this theme has always been central in AVT studies (e.g. Pavesi 2018). In general, it is not possible to expect the same degree of naturalness between a dubbed dialogue and a spontaneous conversation (yet the same consideration applies to the difference between the source text's dialogues – which usually strive to be a faithful representation of spontaneous conversation – and the dubbed version's ones[98]). About prefabricated orality, Hayes (2023) affirms that it can be

> "partly due to the attention […] [dubbing] actors must give to fulfilling lip-sync and synchrony of paralinguistic elements (e.g., sighs, gasps, panting, or laughter) as well as the ideological clash they must over-come when revoicing actors whom they can see are other and whose mouth articulations they can hear belong to another language entirely" (2023: 5).

While the second and the fourth strategies[99] are interesting but not related to what highlighted in Table 4, the third approach that Hayes (2023) exemplifies is the "Foreign-Accent Strategy", which has emerged in recent English-language dubs by Netflix and could be summarised as "the

[98] In particular, Baños and Chaume (2009) show that, at the phonetic and prosodic levels, the domestic (Spanish) productions are much closer to spontaneous conversation than their dubbed counterparts.

[99] The second strategy is the "Domestic-Accent Strategy", where "native […] accents in the original become native-English ones in the dub" (Hayes 2023: 7). The fourth and final strategy is the "Hybrid Accent Strategy", where (most) of "the original actors dub themselves" (2023: 7) in English, to preserve their original accent.

use of foreign accents in English derived from the language of the original" (2023: 7). Two considerations arise from this: first, instances of the foreign-accent strategy can be found in the past as well, in dubs such as the one in *La vita è bella* from 1999; second, even though, this constitutes a creative strategy adopted to avoid traditional standardisation tendencies, in this case, it was perceived as an element that detracts from the naturalness of the scene.

Table 4: Relationship between perceived naturalness and accent in Clip 1

ACCENT (Q1: Did the voices in this clip sound natural to you? […])	
R2	"It is very funny that the main actor is speaking in English but with an Italian accent, that's kind of odd"
R3	"I thought it was weird that the guy talked with an Italian accent […]. For me, it took away from the realness of the clip […]. I don't know, I didn't enjoy it, I would have preferred if it was in a more, you know... my dialect of English or something like that"
R7	"The accents were a bit jarring because, obviously, one of them has like a thick… is it supposed to be an Italian accent? Sometimes it's a little bit off […], it felt a bit strange"
R8	"Not really. It sounded like an Italian person speaking English"

With regards to the perceived naturalness of the dubbed speech in *Clip 2* (the AI-dubbed scene from *Tre uomini e una gamba*) many more remarks were raised. In some cases, criticism concerned similar aspects to those found in *Clip 1*. Some examples are provided by R2 and R3 [12], who made comments on the accent in relation to *Clip 2* as well:

> [12] "They don't sound natural. […] It doesn't really sound like natural spoken English, in the accent that I know of. It sounds like the actors are maybe not native English speakers" (R2).
> "I don't think they're native speakers. They have some kind of inflection" (R3).

These remarks are particularly interesting not only because the accent is confirmed to be a source of impediment to the naturalness of the speech, as in the case of the foreign-accent strategy present in *Clip 1*, but also because the dub of the second clip was generated by speech synthesis software. This is particularly relevant also considering another comment made by R2:

> "The intonation sometimes, the prosody, doesn't sound natural. It sounds like they were trying to stretch out phrases to meet the temporal need to match the length of time to say that phrase, but it wouldn't be said at that speed if it were an actual speech. So, it sounded like they were really trying to force the lines to fit into the different slots" (R2).

The issue of the length of the synthesised speech was indeed mentioned in Section 2.2.1. and investigated, for instance, by Wu et al. (2023), who also talk about naturalness and explain that:

> "as the speech duration of words/characters in different languages varies, the same number of words/characters in the source and target sentences does not guarantee the same length of speech. Therefore, TTS has to adjust the speaking rate of each word in a wide range to match the total speech length, which will affect the fluency and naturalness of synthesized speech" (2023: 13772).

In other words, TTS systems need to carefully balance the speaking rate to achieve a good result in the AD in terms of prosody, rhythm, intonation, and stress patterns. In this case, the speech was perceived as unnaturally slow. However, since the admins of 'italiancomedydub' did not reveal the tools used for the dubs (as already said in Section 2.3.), a detailed analysis of the nature of these problems cannot be provided here.

4.1.2. Audio quality

Respondents' answers to Q1 for *Clip 2* also highlighted issues of different orders. An example is provided by the code AUDIO QUALITY. Indeed, as can be observed in Table 5, the amateurish appearance of the AD in this clip was revealed by a poor management of sounds and volumes of the voices. Indeed, some respondents thoroughly underlined problems in volume balancing (i.e. discrepancies between the voices and the actors' distances, the camera movements, and the setting in general).

Table 5: Relationship between perceived naturalness and audio quality in Clip 2

AUDIO QUALITY (Q1: Did the voices in this clip sound natural to you? If not, can you describe when or why the voices felt particularly unnatural?)	
R2	"I guess it's just a question of audio quality: the voices sound very superimposed. It doesn't sound like they're coming out of the video. It sounds like you're watching a still picture and then you're hearing the speech from outside of it. The depth perception of the sound, it sounds odd"
R3	"The mixing is really off. Like, it sounds like they have one stereo track with the voices and another stereo track with the music […], sometimes the voices are really high, they didn't normalize any of it. So, both the acting and the mixing are terrible"

It is worth to mention that this aspect, i.e. the poor mixing of sounds and voices' volumes in the wider spatial context, was considered as one of the main parameters to affect speech naturalness also in the second clip selected from the 'italiancomedydub' project (i.e. *Clip 4*). The comments in Table 6 testify this and evidence how critical technical tasks are in addition to creative ones to prevent viewers from noticing inconsistencies that can disrupt their film watching experience. In light of this, it can be safely affirmed that, at least as far as amateur productions are concerned, AD still presents considerable technical limits, especially if compared to dubs handled by professionals in dubbing studios. As de los Reyes Lozano (2023) points out, the pace of technological advancement is today more accelerated than ever before. Indeed, "if we think about subtitling or dubbing, what was until recently considered science fiction has now become an astounding reality that is within everyone's reach" (2023: 14). On one side, thus, it is reasonable to think that in a few years technical problems will be easily fixed also in accessible AI-powered AVT solutions for amateurs. On the other, however, volume balancing and general audio depth were now perceived as critical areas of improvement related to the perceived naturalness of the speech.

Table 6: Relationship between perceived naturalness and audio quality in Clip 4

AUDIO QUALITY (Q1: Did the voices in this clip sound natural to you? If not, can you describe when or why the voices felt particularly unnatural?)	
R1	"What I think is missing is the volume, again. Because, at the end, when they are walking, you can hear them at the same volume level, and this is different from original versions when you hear original voices"
R6	"It felt like he was very distant. It felt like he wasn't at the same distance from the camera as it should be. Like, if I'm very close up, you'd expect my voice to be louder. And if I'm further away from the camera, you expect it to be much quieter. It felt like the volume levels weren't quite the same. He was a bit too quiet, a bit, too far away"

4.1.3. Language

A third aspect that is worth mentioning regarding naturalness in the clips selected from 'italiancomedydub' could be encoded as LANGUAGE. This is because many comments focused on a vast array of issues (e.g. non-spontaneous or non-idiomatic expressions, grammatical mistakes) stemming from the intermediate step in the automatic dubbing workflow, i.e. the translation/adaptation step. In a fully automated dubbing, this would imply the use of machine translation tools with little or no human intervention; however, in the specific instance being referenced here, the translation was handled by humans. Indeed, as seen in Section 2.3., the script in the target language is the product of the collaboration between the admins of the Instagram page

and their followers. As Table 7 shows, audience identified the use of language in the target script of *Clip 2* as significant obstacle to the speech naturalness.

Table 7: Relationship between perceived naturalness and language in Clip 2

LANGUAGE (Q1: Did the voices in this clip sound natural to you? If not, can you describe when or why the voices felt particularly unnatural?)	
R3	"There are also some grammatical mistakes. Like, she says 'Plato say', instead of 'Plato says', which is really… that's weird"
R9	"It sounded like it was translated exactly from Italian to English. The first clip sounded more how Americans or native English speakers actually speak, and this one sounded like it was just translated from Italian to English, and it was just read that way"
R10	"It did not seem like a native American English speaker did the translation. From everyone, from all four characters, they were just a lot of words that we wouldn't use in America"

Reception studies often seek to explain the role that AVT and AVT choices play in the film viewing experiences. These comments highlight how a faithful rendering of the source text goes beyond the mere linguistic equivalence, and how a poor translation can have a similar (or even greater) impact on the overall credibility of the dubbed product as the mismatch between the pronounced words and lip movements. Besides grammatical errors (indicative of an amateur process that was evidently not subjected to sufficient revision and supervision), indeed, the correct use of language also involves the challenge of reproducing a natural-sounding dialogue that does not appear forced, i.e. with an appropriate register, style, syntax and lexicon. To enrich the present discussion, the answers to the fifth question [13] in the interview, which sought to ascertain the intelligibility of the dubbed speech across all four segments, could be now mentioned.

> [13] Q5: Overall, were there any moments when the dubbed speech felt unclear or difficult to understand?

For all four scenes, language was defined as generally intelligible by the audience. Nevertheless, some remarks are still interesting to mention here, specifically with regards to C*lip 1*, and *4*. In particular, in *Clip 1*, despite the overall clarity of the language, two respondents identified the Italian accent as a potential source of difficulty in comprehending the scene, as the answer by R7 well summarises: "Well, definitely the thickness of the Italian accent doesn't help, but I don't think I had difficulty understanding it". This confirms that, in some cases, foreign accent strategies

in dubbing could be received as obstacles for a seamless communication. For what concerns *Clip 2* respondents underscored the same concept. With regards to *Clip 4*, answers were aligned to what was previously mentioned, as remarks were directed towards linguistic choices in the target script. What is of particular importance here is that the fact that the translation/adaption was carried out by amateurs is evidenced by the difficulty some respondent had in grasping the cultural reference in the joke that the character Guido makes in the scene [14].

> [14] "The last thing that he was saying as they were walking away, 'Visigoths', maybe it was a made-up word. I'm not sure what word was that supposed to be" (R2).
> "I just think that's a really strange word. I mean, it makes sense in Italian but in English I've never heard anyone say 'Visigoth' in my entire life […]. I mean, probably I would have picked something a little bit more culturally relevant" (R3).

These last observations highlight the relevance of the translation process within the dubbing workflow. As it has been stated on numerous occasions, technologies are now being integrated into professional practice to satisfy the unprecedent demand of translated audiovisual content in the new globalized world. In this context, however, de los Reyes Lozano and Mejías-Climent (2023) underlines how there is a certain resistance to the incorporation of MT engines into AVT due to the well-known constraints of audiovisual texts (i.e. their multimodal nature), and due to some levels of "interpretation and creativity, or hermeneutics, which, for the time being, machines do not seem to be able to deliver" (2023: 4). Nevertheless, results show here that translation done by humans lacking adequate training and skills can be detrimental in the same way.

4.1.4. Other conceptualisations of naturalness

To conclude this part, it is interesting to mention a code that emerged with regards to the naturalness of speech in *Clip 3* (the Netflix-commissioned English dub of a scene from *L'incredibile storia dell'isola delle rose*). It is important to consider that the context of production of *Clip 3* differs significantly from the first and the second clip. If the first excerpt is a quite experimental dub from the late 1990s, and the second is an example of non-professional English dubbing generated through a text-to-speech technology, the third belongs to a new and rapidly consolidating tradition, i.e. that of English-language dubbing on OTT platforms. Contemporary professional dubbing studios and modern production environments, however, are not free from issues concerning their end-products. Indeed, five of the ten respondents did mention an aspect that seemingly disturbed the perceived naturalness of the dub. This aspect could be categorised under the code SPEECH-TO-BODY CORRESPONDENCE. The idea for this code, which summarises the considerations of half of the respondents, is to be linked to the "Speech-and-Sound Rubric" developed by Spiteri

Miggiani (2024) in her quality assessment model of the dubbed speech. In our case, SPEECH-TO-BODY CORRESPONDENCE is to be intended as a macro-label encompassing two of the quality indicators provided by Spiteri Miggiani, i.e. "voice symbiosis" and "body language". In the first indicator, the focus is on the "suitable voice casting according to age, gender identity, physique du rôle, characterization, narrative-related features" (2024: 59). Additionally, the second indicator "refers to the semantic and synchronous correspondence between facial expressions, body gestures, and the uttered target-language speech" (2024: 64). In accordance to this, some respondents addressed body language as the main problem [15], while others focused on inappropriate voices [16].

> [15] "They felt a little bit more natural. Just her, I felt that some expressions or noises that you can clearly see she is making but you don't hear from her... I felt that was a bit unnatural" (R1).

> [16] "Yes, but not for the people that they were assigned" (R8).
> "The girl was fine, but I feel like the guy, he sounded like he was like 18 years old but looked like he was 35" (R9).

Similar considerations regarding naturalness were also raised in the context of *Clip 4* (the AI-dubbed scene from *La vita è bella*). Specifically, SPEECH-TO-BODY CORRESPONDENCE was mentioned by three of the respondents as a significant impediment to the speech naturalness. In this case, it is plausible that the comments were a consequence of the fact that another scene from the same film had already been viewed. Indeed, comparisons were made to the first clip, as well as references to a supposedly different approach to the dubbing of the film [17]. In another case, voice symbiosis was the highlighted aspect [18]. As previously stated with regards to lip synchronisation, speech-to-body correspondence will be touched in greater detail in the next section, entirely dedicated to synchronisation and voice matching.

> [17] "Is it a different version? With different dubbing actors? The voice of the father is kind of funny. It is like a 1920s fast talking detective, it is really old timey the way that he speaks, but I mean the movie is set in the 1940s so maybe that's where they were going for to match like an historical accent, I guess" (R2).

> [18] "Maybe the dad sounded like he was a little bit older than he looked in the film, but it wasn't anything like the last one where you like were like: 'Wow, this sounds like a child's voice and that's a grown man'" (R9).

The latter and the previous remarks may provide insight into the world of AD and voice imitation. In Section 2.2.1., it was shown how this technology enables users to obtain cloned voice features in the audio output. It was further posited that this could potentially radically transform the reception of translated audiovisual texts. Even though this seems impossible to ascertain through a small-scale qualitative study like the present one, negative remarks regarding the naturalness of the voices in *Clip 3* indicate that the casting of suitable (human) dubbing actors is a delicate and crucial step concurring for the quality of the final product. Of course, the utilisation of voice cloning technology in a fully automated dubbing could prevent the issue. However, other potential problems may emerge.

4.2. Synchronisation and voice matching

4.2.1. Lip synchronisation

As stated at length in this thesis synchronisation is the main constraint in film dubbing (e.g. Pavesi 2005). While this particular subsection is actually concerned about lip sync issues in the selected clips, the next one will focus on another type of synchronisation – what Pavesi (2005: 15) defines as "sincronismo paralinguistico" and Spiteri Miggiani (2024: 64) as "body language" – and also on a general coherence between the dubbed voice and the source actor's characteristics in terms of, for instance, physical appearance, age or narrative role.

From the answers analysed so far, it is evident that lip synchronisation is a significant concern among native English-speaking audience, to the extent that in many cases this was the primary aspect that respondents focused on when assessing the degree of naturalness of the speech in the various scenes. However, the second question in our interview [19] was designed to find out to what extent lip sync was a problem for the audience:

> [19] Q2: How well did the dubbing sync with the lip movements of the actors? Did you notice any distracting or unrealistic mismatch?

Whereas previously the results exhibited a certain degree of heterogeneity, potentially due to the open-endedness of the naturalness theme, the answers to Q2 consistently highlight a clear criticism of lip sync in nearly all scenes. Indeed, despite the involvement of professional translators/adaptors and dubbers, *Clip 1* was subjected to considerable criticism. In particular, of the ten respondents, seven noted a poor lip syncing. Some comments can serve to illustrate this point, as one respondent defined lip sync in the scene as "bad […] [and] distracting" (R5), and another one highlighted a temporal disconnect, "especially at the end of sentences […] where the speech is still going but the speaker's mouth starts to close and it doesn't seem like he's actually

saying what you're hearing" (R2). Nevertheless, some positive evaluations were present as well: R3 said "it worked", R4 defined dubbing as "pretty good", and R8 said that "overall it synced pretty well". *Clip 3* also obtained a widespread appreciation: ten out of ten respondents said that the output was better in this case and that "a really good job" (R6) was done.

In contrast, the critical response to *Clip 2* (the restaurant scene from *Tre uomini e una gamba* dubbed with AI) was more homogeneous and pronounced. In this instance, the totality of the respondents agreed in affirming that lip sync was "really off" (R2), "terrible" (R3), "pretty distracting" (R4), or "not good at all" (R5). These remarks can lead to a consideration about priorities. The reception study by Guerberof-Arenas et al. (2024) mentioned in Section 3.1.3. showed that, for what concerns subtitles, translation and language rendering are of great consequence to the overall understanding and enjoyment of the final product. In the case of dubbing (especially if AI-powered), other issues seem to play a more decisive role. As a matter of fact, if translation and language use were identified as major issues to be taken into account in order to guarantee the naturalness of a dubbed film, concurrently it becomes evident that, from the perspective of native English speakers, poor lip sync still draws most of the attention, remaining the main source of concern. Of course, considering new speech synthesis software that enable users to modify the target video's actors' lip movements according to the target script, these considerations might be completely different. It could be thus interesting to study the reception of other types of AI-dubbed products.

In the context of *Clip 2* and *4*, on the other hand, some answers are particularly interesting to mention, as they provide the opportunity to reflect on the concept of multimodality that was introduced in Section 1.1.1.. Indeed, while still affirming that there was no match between the speech and the lip movements in *Clip 2*, R1 and R10 recognised the impact of camera angles in the overall viewing experience [20], i.e. how they can, at the same time, emphasise or mitigate lip sync-related problems. Similar comments emerged in relation to lip synchronisation in *Clip 4*. In particular, while five respondents still stated that lip sync was problematic to some extent, most of the answers focused on the position of the camera, as in those provided by R2 ans R5 [21].

[20] "For me it did not match. Especially at the beginning, when the three men are talking. Because the camera focuses on each one of them and you see them at close. In the second part, since the person who is talking is not necessarily on camera, in that way you get more into it and it is less distracting" (R1)
"Nothing particularly distracting. […] the camera was more focused on the setting and other things happening in the scene, so that could be a factor" (R10)

[21] "In this case a lot of it was them walking away so you don't see their faces very much" (R2)
"It seemed well because they were facing away from the camera" (R5)

In previous sections we discussed the importance of the diverse semiotic modes present in a filmic text, and how they contribute to the construction of its overall meaning. It is clear here that the relationship between – for instance – soundtracks, shots, or set design and the audience's reception of the work is not identical in the original and translated version. In other words, the foreign audience's experience is always different from the original audience's one. As Pavesi (2005: 15) points out, translators/adapters and dubbing actors are not always obliged to respect the original text's synchronisations. This is due to the fact that a film is constituted by a set of semiotic modes, camera work being one of the main ones, and dialogue scenes are not always entirely covered by shots in which the camera is fixed on the characters' faces.

An aspect that is worth mentioning is that considerable attention was paid to mouth movements and to overall gesturing of the Italian actors onscreen by native English speakers who are familiar with the Italian language (because they speak it or they are trying to learn it) such as R1 [22] and R2 [23] and R6 [24]. These examples demonstrate two things. First, as Delabastita (1990) states, "the pronunciation of different languages obviously has a different visual impact" (1990: 99). Second, sociocultural variables always influence the reception of a work. Indeed, despite the widely held belief that Anglophone countries tend to have a strong resistance to dubbing, it is important to note that "audiences are also revealing a different tolerance threshold in terms of, for example, the accuracy of lip synchrony in dubbing" (Chaume 2018: 51). This is certainly linked to sociological and individual variables, with "social class, educational background and the individual's own command of a foreign language proving to be important factors influencing the strength of the preference stated" between dubbing and subtitling (Herbst 1997: 291).

> [22] "Not just the lip movements but also the overall gestures of him talking […]. The body language compared to English… if felt that something was missing. Probably this is because I know how Italians move when they speak" (R1 on lip sync in *Clip 1*).

> [23] "I mean, it's always something that's never really perfect. It wasn't that noticeable here, but of course, sounds in English and sounds in Italian are quite different. English words end in consonants, Italian usually end in vowels. So like, the word boundaries usually tell there's something a little bit off" (R2 on lip sync in *Clip 3*).

> [24] "I don't think it lined up very well at all, but I don't think it mattered too much, because obviously there's a lot more syllables, I think, a lot more mouth movements in Italian. So, you could tell that the mouth was moving quite a lot, and he was speaking quite quickly, which you don't really need to do in English, not to the same extent anyway" (R6 on lip sync in *Clip 1*).

To conclude this part, it is important to make a reflection. As Gambier (2018) explains, when actual translated films viewers' opinions are elicited, "the risk is that the perception and reception is limited to the micro-level, forgetting larger cultural factors" (2018: 48). On one side, overall, results confirm lip sync as the main constraint in film dubbing. On the other, if lip sync appears to be the main concern in our sample (composed of young, educated people who speak foreign languages and prefer subtitles to dubbing in the totality of cases), answers could be of a different nature if other subjects with different backgrounds were interviewed (such as Italian viewers, who are typically more familiar with this kind of disconnect and may find it less disturbing).

4.2.2. Speech-to-body correspondence

In Section 4.1 the code SPEECH-TO-BODY CORRESPONDENCE (STBC) was identified as one of the main issues regarding the lack of naturalness in dubbed speech. It is useful to reiterate that when the expression STBC is employed, it should be understood as encompassing both voice matching (or symbiosis) and body language, i.e. "the appropriate attribution of voice qualities to speakers" (Spiteri Miggiani 2024: 64) in the target audio, and "semantic correspondence and synchrony" (2024: 60) between speech and facial expressions/body gestures.

For its importance, this code is here considered as a distinct theme within the present thematic analysis. Indeed, the third question [25] in the interview concerned this aspect, aiming at investigating, ultimately, the interplay between images and dubbed dialogue. The reasons behind this question are of two different orders. First, as Pavesi (2005) underlines, "sarebbe un errore circoscrivere l'importanza del sincronismo alla sola articolazione labiale" (2005: 13). Lip movements, thus, are only some of the many other elements that need to be handled when dubbing an audiovisual text. Indeed, paralinguistic (or expressive) synchronisation (2005: 15), i.e. the connection between speech, gestures, body and facial movements, is perhaps of the same importance. The second reason concerns voice symbiosis. As elucidated in Section 2.2.1., new speech synthesis models are capable of cloning the original actors' voice features (e.g. Yang et al. 2020). In the context of the videos selected from 'italiancomedydub', the task of voice imitation is present. Nevertheless, in light of the respondents' lack of familiarity with the original voices of the Italian actors, it was interesting to verify to what extent voice imitation is sufficient to ensure STBC.

> [25] Q3: Do you think the dubbed voices matched the physical appearance and behaviours of the actors on screen?

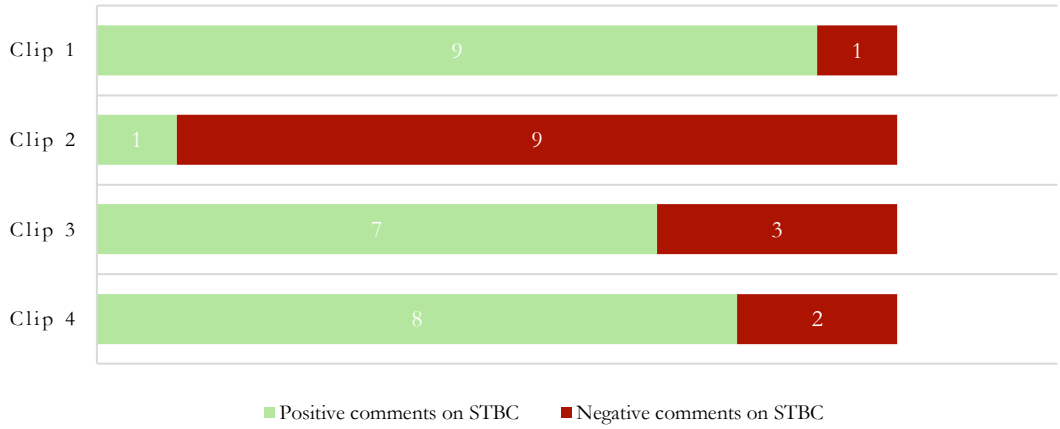
Figure 3 graphically summarises the answers to Q3. For the most part, respondents highlighted a good STBC. This is true for the two human-dubbed clips, as well as for one of the

two AI-generated dubs (*Clip 4*). More precisely, the first clip was very well received, as underlined by R6 ("the main actor, I think he did a fantastic job, I think it was very believable"). With regards to the third clip, the code of VOICE SYMBIOSIS could be identified, as three respondents mentioned aspects related to this as the main problem. Criticism was leveled at the male's protagonist voice (R7: "Giorgio, I felt like his voice seemed too nasal and kind of high-pitched") and at the police offer's voice (R5: "It was really difficult to imagine a police officer having a voice that was so, like, mellow and high"). Concerning this aspect, Bosseaux (2019) explains that "although there are studies positing that it is important to maintain the qualities of the voices of the original actors […], more reception studies are needed to ascertain what the actual impact of voice attribution is on dubbed products" (2019: 54).

In a recent article, Sánchez-Mompeán (2023) sought to identify the reasons behind the relatively low acceptance of dubbed films by English-speaking audiences. Particular emphasis was attributed to the voice symbiosis, as "complete unity between the voice we hear and the body we see on screen is of utmost importance for credibility and authenticity" (2023: 9). Negative comments towards STBC could be explained, in this context, by two considerations. First, due to the absence of a well-established dubbing tradition in Anglophone countries, voice casting and overall quality in the dubbing actors' performances might not be excellent. A second more viewer-oriented explanation regards English audience's lack of familiarity with dubbing as an AVT method. Indeed, "the deliberate efforts made by Netflix to assess and enhance the quality of English dubbed versions could prove insufficient as long as audiences do not fully familiarize themselves with the dubbing mode" (2023: 12). Thus, since in many cases native English-speaking viewers are reluctant or exposed to dubbed content for the first time, their suspension of disbelief[100] is not comparable to that of audiences who are culturally used to hear dubbed speech and have a higher tolerance threshold. In this context, Sánchez-Mompeán (2023) suggests that these issues could be solved and that "continuous and long-term exposure to dubbed content might be an essential step in attaining this goal" (2023: 12).

[100] The expression, coined in 1817 by the poet Samuel Taylor Coleridge, refers to the reader's (and viewer's) willingness to accept the limitations of fiction for the sake of its enjoyment. Indeed, "even if spectators are fully aware that what they are watching is not real, they can still […] become completely absorbed by it thanks to a series of cognitive processes stimulating immersion […] which can vary according to the fictional content and the spectatorship" (Sánchez-Mompeán 2023: 5).

Figure 3: Speech-to-body correspondence reception in the four clips



The red bars indicate instances of negative comments regarding SPEECH-TO-BODY CORRESPONDENCE in the video clips. As it can be observed, *Clip 2* was widely criticised. Specifically, while some comments focused on the mismatch between speech and gestures (R2: "the gestures are very Italian gestures and that's kind of hard to see that in an English speaking context, because I don't know what you could say in English that would match the gestures that they were making"), the majority of the respondents addressed to the issue of VOICE SYMBIOSIS, as Table 8 summarises. As anticipated when discussing about the concept of naturalness, the fact that an AI-generated speech using voice cloning technology elicits such negative comments is an aspect worthy of further investigation. Indeed, this might be related to diarization, i.e. a process in automatic speech recognition when an audio input is automatically divided into segments according to the identity of the speaker. In relation to this, Yang et al. (2020) stated: "we have found that in complex videos with rapidly alternating speech […], diarization […] is insufficiently accurate for decoding who the active speaker is. Future work is required in this direction" (2020: 14). It is reasonable to assert that, in this case, AD had a detrimental effect on the characterisation of the characters.

Table 8: Voice symbiosis in Clip 2

VOICE SYMBIOSIS (Q3: Do you think the dubbed voices matched the physical appearance and behaviours of the actors on screen?)	
R1	"When the three of them are talking, there is nothing distinctive. They all seemed to have the same voice"
R3	"They sound totally out of place. This guy's got a mustache. He looks like he's 50. The voice actor sounds like he's in his 20s, you know He sounds like a young buck"

R5	"I felt like all three of the men had the same voice"
R6	"It was actually kind of difficult for me to tell who was supposed to be talking. The male voice, at least, the main two sounded very similar"

Figure 3 shows an interesting detail. In contrast to the *Clip 2*, the second AI-generated dub (*Clip 4*) appears to maintain a good degree of STBC. Since respondents had not been previously informed that the clip contained synthesised voices, it is reasonable to posit that their answers were influenced by the assumption that the voices were the same as those in *Clip 1*. If on the one hand this could be a sign of a good output, on the other it could also highlight a certain inconsistency with regards to AD. Indeed, the reception of the two clips containing synthesised speech was almost completely opposite and voice cloning did not seem to be enough to ensure good levels of voice symbiosis.

4.3. Emotional tone

As mentioned in Section 2.2., AD technology is already being used for translating audiovisual content belonging to textual genres such as news broadcasting, which are tendentially less constrained by the many limitations that filmic texts display. Indeed, as Georgakopoulou (2019) underlines, a "factor to be taken into account when considering audience reception of synthetic speech is […] the genre of the video material […], as some productions might be more acceptable with 'flat' synthetic voices […] than others" (2019: 529). The fourth question in the interview [26] aimed at investigating, albeit indirectly, to what extent AD can be used in fiction, a genre that requires high levels of rendering of the source actors' emotional tone.

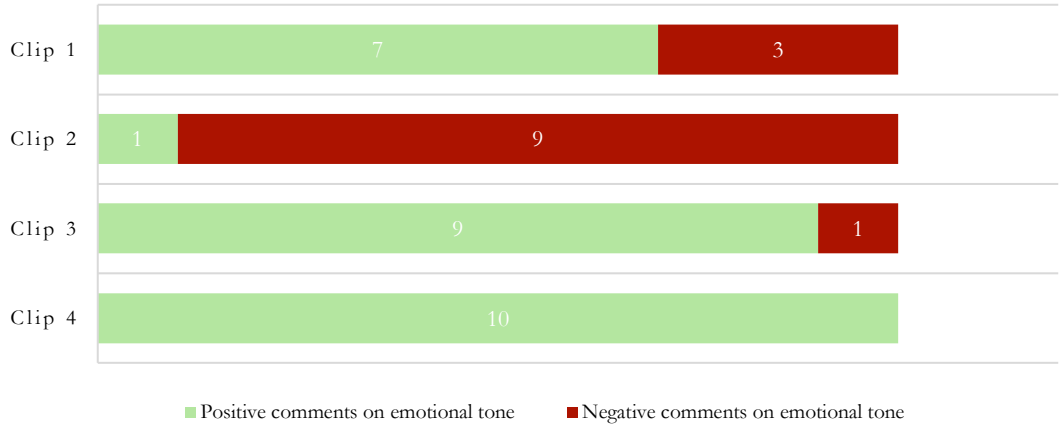
> [26] Q4: Do you think that the dubbed version managed to capture the original emotional tone of the scene?

As previously stated, on the one hand it is true that "viewers play a significant role in making dubbing work" (Sánchez-Mompeán 2023: 12). This implies that dubbing can work properly only if viewers accept the fictional nature of what they are watching and become fully immersed in the narrative. On the other hand, audience's unfamiliarity cannot be used as an excuse for the poor quality of a dubbed product. In this context, emotional tone is a theme of utmost importance and requires careful examination. Emotional tone in dubbed speech has been analysed in AVT studies

through different lenses. For instance, Naranjo (2021) investigated different dubbing styles in term of emotional tone. In the study, the preferences of Spanish viewers were examined through quantitative and qualitative methods. Results revealed that viewers do not *a priori* prefer a specific dubbing style (e.g. natural voices aiming to sound as spontaneous as possible vs. play-acted voices), but rather accept what is considered to be more suitable for the nature of the scene.

As said in Section 2.2.1, one of the primary challenges of synthesised voices is the accurate reproduction of the emotional nuances present in the performances of the original actors. While professional studios might already be capable of achieving this, amateurs engaged in cyberdubbing practices still seem to lack consistency in this regard. Figure 4 visually summarises the comments emerged in relation to Q4.

Figure 4: Emotional tone reception in the four clips



Although the choice of dubbing the protagonist (Benigni) using an Italian accent was criticised – as seen in Section 4.1. – the emotional tone of *Clip 1* was well received by seven of the ten respondents. The scene, which could be categorised as 'tragicomic', because humorous and dramatic at the same time, managed to evoke the same sense of humour and sadness in the audience through the English dub (although three respondents commented negatively on the dubbing actor's performance). What is interesting is that *Clip 4* (the AI-dubbed scene from *La vita è bella*) received positive comments only, and the emotional tone was defined as "the same as the first one" (R3) or even "a little bit better" (R2). One hypothesis that may be posited in order to explain the nature of such comments is that respondents did not perceive any distinction between the two versions (i.e. *Clip 1* and *2*), as testified by R9's answer: "I mean, I knew it was from the first clip, so I felt sad again seeing it […], it has that same emotional tone". This also suggests that, in this instance, the synthesised vocal output was of a high quality.

Even though respondents were not aware of the artificiality of the dubbing in *Clip 2*, the emotional tone of the clip was generally more criticised, with R6 even suggesting the possibility of

an AI-powered dubbing ("It felt like it was an AI dubbing over, like there wasn't any emotion in the words that they were saying"). A code could be identified within the nine answers commenting the dubbed speech in a negative way. The code could be summaries as DIRECTION, since three respondents demonstrated varying degrees of appreciation towards the emotional tone in relation to different parts of the scene. As illustrated by Table 9, respondents clearly intercepted these two distinct moments in the clip. Overall, inconsistencies in the emotional tone could be interpreted as signs of the influence of extra-linguistic elements present in the scene, which pertain to the soundtrack, to the photography, and, more broadly, to the direction of the scene. Indeed, as anticipated in Section 3.2.1.2., the restaurant scene could be seen as divided into two parts, the first more static with shots closer to the characters, and the second with wider camera movements and music in the background.

Table 9: Relationship between direction and emotional tone in Clip 2

DIRECTION (Q4: Do you think that the dubbed version managed to capture the original emotional tone of the scene?)	
R1	"At the end maybe, it did. Because at the end there are less close-ups. You don't get distracted by the dubbing. So maybe, at the end it sort of captured the humour. At the beginning it was more difficult to get into it"
R3	"Maybe towards the end of the scene"
R10	"Second half, yes, with the woman. First half, no"

It is essential to recall the key points discussed in Section 1.1.1. regarding the concept of multimodality. Indeed, it was highlighted how "AV implies quite a number of signifying codes that operate simultaneously in the production of meaning" (Gambier 2018: 50). Accordingly, semiotic modes can influence the reception of AVT. This is confirmed by Naranjo (2021) who states that the presence of background music can play a central part in influencing emotional tone perception (2021: 595). Whereas all these factors are routinely taken care of by audiovisual translators in studios and by dubbing actors in recording booths, automatic dubbing does not seem, as far as the results in this study show, to be able to ensure uniformity in emotional tone under all conditions. Indeed, "TTS models that rely purely on text scripts […] lack style information and tend to generate neutral waveform, which cannot meet the requirements of vivid dubbing" (Liu et al. 2024: 158).

Many are the studies that propose solutions to this issue. An example is provided by Liu et al. (2024) who developed a multimodal speech synthesis method which enhances the expressiveness of the synthesised audio by extracting features from the source video.

4.4. Attitudes and opinions

4.4.1. Identification task

Before proceeding with the thematic analysis of the second set of questions (which involves opinions and attitudes towards the use of AI in the field of AVT), it seems now pertinent to consider an additional dimension of the interview. Indeed, as anticipated in Section 3.2.2.2., the interview was structured as follows: after the questions about the reception of the four clips, it was revealed to the subjects that two of the clips had been dubbed using speech synthesis technology. Subsequently, respondents were presented with an identification task. Specifically, subjects were asked to identify (based on their memory, i.e. without looking at the clips again) the two clips generated with AI. The identification task was used with the intention to obtain a more nuanced perspective on the reception of the dubbed clips by the audience, by capturing patterns within the observations expressed by the interviewees. Before analysing the results, it is worth reiterating now that *Clip 2* and *4* were those realised by the admins (and followers) of the 'italiancomedydub' project, while the other two clips (*1* and *3*) were examples of English-language dubbing made in professional contexts by human experts.

Results, which are graphically represented below in Figure 5 and Table 10, show interesting data. First, Figure 5 immediately highlights that no clip 'scored' ten out of ten, i.e. there was never a unanimous agreement in the correct categorisation of the clips. Interestingly, the majority of respondents (eight out of ten) thought that *Clip 3* had been created thanks to AI-powered tools. At the same time, it is noteworthy that the two clips from *La vita è bella* obtained the exact same score, i.e. seven respondents thought that both clips were made by humans (although this was not always the case). This finding is consistent with the results of the previous thematic analysis, where parallel considerations about speech-to-body correspondence and emotional tone were formulated for both clips, implying that most of the interviewees assumed that both clips belonged to the same version of the film. In contrast, *Clip 2* was perceived as the most ambiguous, with a more balanced distribution of guesses. Six respondents indicated that they believed it was AI-generated, while four suggested that it was created and dubbed by humans.

Figure 5: Identification task

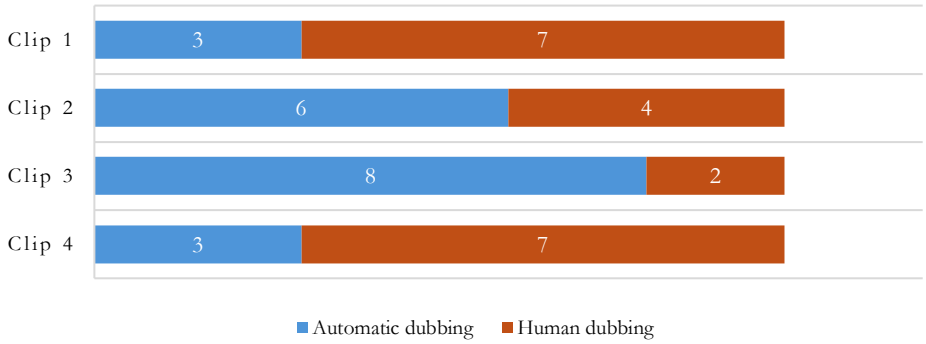


Table 10: Individual guesses in the identification task

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
Clip 1	AD	AD	HD	HD	HD	HD	HD	HD	AD	HD
Clip 2	HD	HD	AD	AD	AD	AD	HD	AD	HD	AD
Clip 3	HD	AD	AD	AD	AD	HD	AD	AD	AD	AD
Clip 4	AD	HD	HD	HD	HD	HD	AD	HD	HD	HD

It is not possible to gain a full understanding of the guesses made, without analysing the content of the answers provided. Table 10 demonstrates the aforementioned lack of consensus on the nature of the clips, i.e. automatically dubbed (AD) vs. human-dubbed (HD). In this context, the interview format allows to uncover valuable context. Indeed, since the interview consists of open-ended questions and answers, the interviewees could not only make their attempt, but also justify it. Within the responses collected in relation to the identification task, a significant pattern emerged, which could be summarised by the code 'EXPECTATIONS'. As a matter of fact, respondents classified *Clip 3* as the most probable to be produced with the use AI tools, because its perceived quality was the highest. As Table 11 summarises, thus, native English-speaking audience has, in some cases, higher expectations for artificial intelligence than for humans working in the field of audiovisual translation. As reiterated in other sections, this can be linked to sociocultural implicit norms and habits. Indeed, regardless of the recent countertrend promoted by subscription-video-on-demand platforms such as Netflix, Anglophone viewers still appear to be highly sceptical about dubbing.

Table 11: Relationship between expectations and guesses in the identification task

EXPECTATIONS (Q9: Two out of these four clips were made using AI-based tools. Can you identify them?)	
R3	"The reason I said the third one is because the female's voice sounded too realistic. I almost thought, since it's hyper realistic, maybe that's a voice model that almost is too good at. That was my theory"

R7	"I would say the *Rose Island*, I would have to guess that that was AI […]. My guess is just based on the fact that I've used sort of speech synthesis tools […] and I know how accurate they can be in reproducing speech. I guess I considered them to be the most accurate […]. It must have been some really bad AI that did that. As I said, I think it's a very powerful tool, but that doesn't necessarily mean that it's always used in the right way"
R8	"I thought in the first and the third […] the translation was very clear, they sounded more like how a mother tongue English speaker would speak. And the second and the fourth sounded more like someone who wasn't a native speaker would translate them. So that's why I thought that those were AI. I guess I just assumed that AI would be like the best of the best"

R8's answer serves to reiterate the significant impact that a poor translation can have on the overall reception of an audiovisual product. For the clips produced by the admins and users of the 'italiancomedydub' project, in particular, the poor linguistic outcome was highlighted in several occasions. Nevertheless, while this demonstrates the impossibility of directly comparing the output of amateur creators with that of professionals, it is interesting to note that there are higher expectations placed on artificial intelligence than on human professionals. R7's response, on the other hand, with an emphasis on 'the right way to use AI tools', leads us directly to the final questions of the interview, about the opinions and general judgements towards the utilisation of AI in the domain of creativity (and, specifically, in AVT).

4.4.2. Concluding remarks on the reception study

Studying actual audiences' reception in the age of the "ever increasing 'multimedia–isation' of the world" (Taylor 2012: 13) is interesting for a number of reasons. First, it can be helpful to assess the actual quality of specific translated content. In addition to this, reception studies can shed light on preconceptions and consequences related to the utilisation of given AVT methods. In this case, the interview-based study conducted here helped to highlight features of AI-powered cyberdubbed content, as well as characteristics of English-language dubbing in general.

Overall, the initial answers to the interview showed how the perception of speech naturalness varies widely across the audience, encompassing multiple conceptualisations and ranging from audio quality to language used. Results also showed that the themes of speech-to-body correspondence and emotional tone were treated in a similar manner by the interviewees. As a matter of fact, while the majority of the comments towards *Clip 1*, *3*, and *4* were favourable, considerable criticisms was directed towards *Clip 2* for both themes. Finally, lip synchronisation was criticised in all cases. Considering what was outlined in Section 2.2.1. (i.e. rapid advancements in the field of speech synthesis), it is reasonable to hypothesise that the identified issues, particularly those pertaining to language, emotional expression, and technical management of sound in the output, will be promptly limited soon.

In summary, the thematic analysis and the assumptions behind the guesses to the identification task allow to draw two main findings: first, the English-speaking audience's alleged unfamiliarity with dubbing was confirmed, as traditional issues (i.e. lip sync) were highlighted and mentioned as problematic in both AI-powered amateurs dubs and professionally produced dubs; secondly, interviewees exhibited higher expectations for the quality of AI-generated dubbing than for that of human-generated dubbing (as very often AI versions were not identified as such by the audience). Since it is evident that the results are contingent upon the sociocultural background of the respondents, it would be interesting to conduct further research to confirm these data through larger-scale quantitative analysis. Nevertheless, a constellation of observations emerged through the interview. It appears finally possible to state that most of the data collected could not have been elicited through pre-packaged and close-ended questions.

4.4.3. What futures for audiovisual translation?

The present thesis has previously addressed the distinction between subtitling and dubbing, and their global dissemination in accordance with audience preferences. The analysis of the answers has confirmed the supposed preferences, i.e. native English-speaking respondents revealed that they have a strong preference for subtitling over dubbed content when viewing foreign-language productions. Dubbing is thus often seen as something that interferes with the cinematic experience, although "for many cinema professionals and film buffs, subtitles are a blemish" (Díaz-Cintas and Remael 2007: 82) as well. As already specified, this thesis does not aim at establishing the best AVT method. Instead, following this confirmation of the audience's preferences, the final phase of the interview aimed to investigate potential future scenarios in the AVT, i.e. whether AI can contribute to expanding the scope of dubbing and whether its use in the AVT field will be widely accepted. In other words, after the confirmation of today's habits, an attempt was made to identify what tomorrow's possible habits might be. This was done following Gambier's concept of "repercussion", which deals with preconceptions and consequences related to a given AVT mode (Gambier 2018: 57). The final questions of the interview listed below were elaborated to gain a deeper understanding of audience's opinions and attitudes towards general artificial intelligence [27], automatic dubbing [28], and future trajectories in the field of AVT and entertainment [29].

[27] Q6: "What is your opinion on artificial intelligence?"

[28] Q10: "How do you think knowing in advance that some of these clips were made using AI would have affected your reception?"

[29] Q11: "Would you be willing to watch AI-dubbed films in the future?"

It is clear that each of these questions brings forth a range of ethical considerations within their respective answers. Q6, in particular, triggered long and elaborate answers. Therefore, it would be impossible to create a graph that clearly distinguishes between positive and negative comments on the topic in question, as has been previously done. Nevertheless, patterns can be identified. Specifically, two types of comments emerged. First, some responses demonstrate a good awareness and knowledge regarding the nature of AI [30]. This may be attributable, in some way, to the composition of the sample, which consists precisely of young, computer-literate and educated individuals. Another identifiable pattern is the attention placed by respondents on the conscious use of AI tools. Indeed, as the remarks by R1, R4, R5 and R6 [31] well highlight, it can be said that the general attitude is mixed. On the one hand, there is excitement and curiosity about AI in general. On the other, respondents show a certain concern, not about the features offered by the tools themselves, but about the unethical use that humans can make of this technology.

> [30] "I think it's just a tool, you know, it's mostly based on stochastic learning. I think when people are scared about it it's just because they don't know very much about it, but it's not really intelligence, you know? […] That's just really good probability" (R3)
> "Generally, when I think of artificial intelligence, I think of generative AI in the context of large-language models like ChatGPT. So, I think of something that can read and analyze a lot of texts very quickly, much faster than any human can. I don't think it's sentient, it can't think for itself, but depending on your input, and what you ask or tell the system to do, then your output will be determined by that. It's a great tool. I love it" (R10)

> [31] "I think it's an amazing tool. I don't find it scary […]. What I am more scared of is how you use it. AI alone, I don't think... It's more us, humans" (R1)
> "It could be useful in some cases but it's also worrying, the way people might use it" (R4)
> "I find it interesting. I think artificial intelligence is a tool. Instead of being scared of it, we should all just learn more about it and learn how to manipulate it" (R5)
> "I find it fascinating and terrifying. I think it can be a fantastic tool to use, but we have to use it, ethically. It can be very easy to misuse" (R6)

In *AI4People* (2018), the group of scholars guided by digital ethics philosopher Floridi focused on the notion of 'human agency'. In relation to the revolution prompted by AI, it was affirmed that: "put at the service of human intelligence, such a resource can hugely enhance human agency […]. In this sense of 'Augmented Intelligence', AI could be compared to the impact that engines have had on our lives" (2018: 692). The last answers above confirm this focus: rather than on AI models *per se*, attention should be paid to human agency, i.e. "what we can do" (Floridi et al. 2018:

690) with AI in order to move forward with the growth of technology as active and conscious members of the society. In the field of audiovisual translation, for instance, Georgakopoulou (2019) points out how new technologies has always pushed the entertainment industry forward (2019: 517). Besides all technological inventions of the 20th century, Georgakopoulou gives the example of the broadband Internet access, which not only "catalyzed the transformation of the translation industry into the globalized and centralized localization industry we know today", but also "had a direct impact on the amount of video content that could be available to consumers, unleashing unprecedented volumes of material" (2019: 521). This well shows the strong relationship between AVT and technology, since "the very existence of AVT is a byproduct of developments in film, video and broadcasting technologies" (2019: 515). In this context, it is impossible not to question the potential impact of AI in AVT.

The graphs below graphically summarise the answers to Q7 and Q8. As it can be observed, most respondents are aware of the current potential of AI in AVT (Figure 6), but the situation is balanced when respondents are asked whether they have ever encountered audiovisual content translated with AI-powered tools (Figure 7), as often respondents state that it's difficult to ascertain whether a certain kind of translated audiovisual material is AI-generated or not [32].

Figure 6: answers to Q7

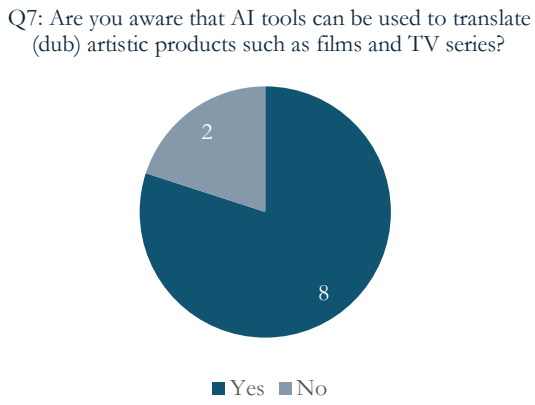
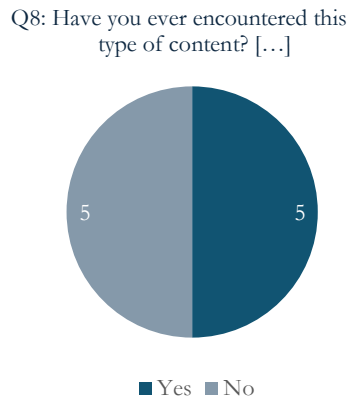


Figure 7: answers to Q8



[32] "In some cases it's almost indistinguishable from reality" (R7)
"I know that a lot of content online is AI generated now, but I don't know to what extent, sometimes I don't know if something is AI generated or not. I don't know what's real or fake anymore" (R10)

Q10 ("How do you think knowing in advance that some of these clips were made using AI would have affected your reception?") leads us back to the discussion in the previous section on expectations, as the results of this part of the interview are consistent with those of the identification task. Indeed, in some cases, expectations for AI were higher than for human

translators/adapters and dubbers. Accordingly, some responses to Q10 seem to reflect the same aspect. With this regard, answers by R2 and R8 [33] provide us with an example of the general English native speakers' scepticism towards dubbing and trust in new technologies.

> [33] "I would have had higher expectations on AI, because I've heard AI generated voices […] and usually they're very very good. So, I'm surprised that it's not that good. I did not identify them because I thought they would have been the better ones, because of what I've heard of speech synthesis and because I've seen some things on Netflix, like Korean series dubbed in English, and it's usually really unrealistic" (R2)
> "I think it would have made me more critical because if it's a computer it shouldn't make mistakes, right?" (R8)

In parallel, another trend that could be observed is related to a much-investigated topic in AVT studies, that of immersion, i.e. "the experience of a viewer or reader of becoming lost in a fictional reality" (Kruger and Doherty 2018: 95). Indeed, some respondents [34] stated that knowing in advance that they are watching content translated using AI might could potentially make them more suspicious and attentive to the quality of the final product, with potential detrimental consequences on the overall immersion. Kruger and Doherty (2018) focused on this topic, providing examples of reception studies that aimed at investigating, though a variety of measures, how immersion can be affected by AVT modes. For example, subtitles were described as particularly impactful. Nevertheless, a study by Wissmath et al. (2009) demonstrated that dubbing is not always more immersive than subtitling.

> [34] "It probably would have put me more on guard" (R3)
> "Probably I'd be thinking about it the entire time watching" (R4)
> "I think I would have been more focused on the quality" (R5)

Being suspicious (or frightened) of automatically dubbed content, however, does not mean being unwilling to experience it. The very last question of the interview (Q11: "Would you be willing to watch AI-dubbed films in the future?") provided insight into the interviewees' preferences and attitudes towards automatic dubbing. In particular, a widespread curiosity towards the topic could be observed. While the theme of conscious use was reiterated [35], most respondents showed a willingness to watch artistic products where translation is fully handled by AI [36]. Therefore, AD is not seen as being inherently negative and detracting from the film-watching experience. Indeed, even if the outputs generated by the 'italiancomedydub' project were not always satisfying in terms of perceived naturalness or speech-to-body correspondence, nine of

the ten respondents declared that they would be open to watch entirely AI-dubbed products in the future.

> [35] "I don't know if I would, because then ethically it's like wishy-washy" (R4)
> "I think the quality is going to get much better, but I don't know if it would be worth it" (R6)
> "I would, because I want to see where things are going […]. I don't want that to become the norm but I also don't want to live under a rock. I don't want to be unexposed to what's happing in the world around me, but I definitely don't like where it's going" (R10)

> [36] "I feel like, as long as you get within what a human would consider to be their perception, then it's fine. If you can't tell the difference between a robot and a human then, for me, it's not a problem. If you can dub an entire film and save money" (R3)
> "Yeah, I'd definitely be willing to watch an AI dubbed film, regardless of the clips. Especially as I think you could have more powerful tools such as choosing whether you want them to speak in the original accent or in an English accent, or an American accent or Australian, whatever. I feel like to have AI features built into streaming services […] to basically tailor the experience to needs. You can't do that with human dubbing because, you know, you do it once and that's it" (R7)

In the previous sections, English-language dubbing was compared with cyberdubbing practices, as both are relatively new and lack a deeply rooted tradition. This lack of established conventions often makes them open to experimentation with creative, non-mainstream solutions. Talking about fansubbers (yet the discourse can be extended to fan AVT in general), Dwyer (2019), affirmed that they "defy the media industry's cultural and language hierarchies" (2019: 443). Another parallel could be drawn between cyberdubbing and automatic dubbing practices. As seen in previous sections, AI-assisted translation is essentially new and experimental as well. As a consequence, it opens up new scenarios in the field of entertainment. From an industry perspective, the advent of these technologies has the potential to transform the way content such as films or TV series is translated and localized (due to the emergence of techniques such as voice cloning and the modification of actors' facial expressions and movements). From a user perspective, AI could revolutionise the way content is consumed as well. As a matter of fact, R7's answer confirms this aspect, as it mentions a possible do-it-yourself film watching experience in the near future. In other words, AI tools may facilitate enhanced personalization options in terms of dubbing, tailoring the viewers' experience to their preferences. This could entail, for instance, the real time control over a range of voices, accents, or even languages. Although these are merely hypotheses at the moment, it is evident that AI has the potential to revolutionise the media landscape, establishing new norms and avenues (but also risks) for both industry players and audiences, as cybertranslation and

English-language dubbing are currently already doing. Such parallels should not be intentended as distinct and separate audiovisual translation methods. Indeed, as the videos realised by the 'italiancomedydub' project demonstrate, English-language dubbing, automatic dubbing and cyberdubbing can cohexist. It is therefore reasonable to conclude that the use of AI, even in professional contexts, will become more frequent (given the vast quantity of audiovisual material that is produced on a daily basis and requires rapid global distribution).

Since this thesis did not address this topic directly, ethical implications of AI in the field of AVT remain unexplored. As Floridi et al. (2018) clearly underlined: "that AI will have a major impact on society is no longer in question" (2018: 690). What is left to explore, at the moment and in future research, is to what extent its utilisation within the field of AVT can constitute an opportunity, without ignoring the challenges associated with the technology noted in Section 2.1.1. (deepfakes, copyright issues, job losses, creative intent, quality standards, and so on). Of course, human objectives must be always taken into account, because there is a bounded range of output when using AI. This means that, in principle, an AI tool produces outputs based on what the human user prompts. In this sense, it seems reasonable to consider that technology will not replace humans unless humans want it. If advanced technologies are used in contexts where they should not be used, technologies are not to be blamed, but humans. The present study revealed that audiences are concerned about the correct use of AI but do not oppose its utilisation *a priori*; rather, they are intrigued and receptive to witnessing future developments.

CONCLUSIONS

The present thesis explored latest trends and future perspectives in the field of audiovisual translation. Trends and patterns have been identified within the answers provided by the interviewees during the reception study. Such answers helped to shed light on three emerging and underinvestigated phenomena: English-language dubbing, automatic dubbing, and cyberdubbing. The results of this study demonstrate that cyberdubbing practices are comparable to those of fansubbing (and cybersubtitling in general), a much more inspected area in AVT studies. Indeed, cyberdubbers defy conventional norms, adopting collaborative and non-hierarchical organisations to produce dubs for online distribution. In addition to this, they deviate from mainstream circuits by experimenting with language (e.g. producing English-language dubs) and technology (using AI-based tools for creative purposes). A second significant finding of the study concerns Anglophone audience's reception of dubbed content: answers revealed that respondents had greater expectations towards AI tools than towards humans, confirming their alleged unfamiliarity with the AVT mode in question and suggesting, at the same time, a great confidence towards artificial intelligence. Indeed, despite widespread concerns about the integration of AI into the realm of AVT, interviewees expressed curiosity about this technology's potential, as all participants stated to be open to watching an entirely AI-dubbed film in the future.

With regards to RQ 1 ("What is the reception of native English-speaking audiences to AI-based English-language dubbing?") it is necessary to mention two aspects. In the first place, the considerations made in the reception sections are confirmed. Indeed, answers indicate that audiences do play an active role in the comprehension of audiovisual texts. Despite the limited scope of the sample, the responses and themes that emerged were numerous and diverse, suggesting that the texts (with their associated AVT methods) are not pre-determined entities to all individuals. Instead, they seem to be susceptible to variation as a result of individual variables. In the second place, reception of automatic dubbing was overall negative. Respondents spontaneously identified typical challenges of traditional dubbing as particularly detrimental for the dubs' naturalness. For example, it emerged that Anglophone viewers consider the lack of lip synchronisation as one of the main impediments to speech naturalness. Interestingly, this applied to both human dubs and AI-dubbed clips. Therefore, it would appear that audiences unfamiliar with dubbing tend to concentrate on a number of 'traditional' issues, irrespective of whether the dubbing is carried out by human or artificial means. As mentioned above, another interesting aspect emerged from the thematic analysis of the answers is that, within the sample, there was a greater degree of confidence in the capabilities of new technologies then in those of English-language dubbing professionals. This lack of trust was especially evident in the answers of the identification task, as many respondents thought that clips presenting human dubbing were actually made using

AI (and *vice versa*). As with the numerous criticisms of lip sync in the clips, these guesses can be explained by the sociolinguistic and cultural background of the sample. The Anglophone audience has, indeed, little experience with English dubbing, which often still suffers from low quality even at professional levels due to the lack of an established tradition. This well demonstrates that, similarly to AD and cyberdubbing, English-language dubbing is also essentially new and needs improvement. These three phenomena could be seen as related to each other, since share the property of being essentially a novelty in the media landscape. Moreover, lacking an established tradition and emerged as consequences of the modern digital world, all three practices exhibit high degrees of creativity and experimentation.

In relation to RQ 2 ("How do cyberdubbing practices differ from those of professional mainstream dubbing?") it should be noted that the objective of the interview was not to identify which type of dubbing was superior. Such a comparison would have led to predictable results and would have been unnecessary. The two types of dubbing were juxtaposed in order to discover, in parallel, their respective features and to elicit general opinions and attitudes towards them. While similar results emerged for the first, third and fourth clip, *Clip 2* (AI-generated cyberdub of a scene from *Tre uomini e una gamba*) was the subject of most criticism in terms of perceived naturalness, emotional tone, and speech-to-body correspondence. Films are complex semiotic systems where language inevitably play a crucial role in the construction of the overall meaning and value. An interesting aspect to mention, which really marks the difference between clips created by amateurs with AI and those produced by human professionals, is the use of language. Indeed, even though the clips were not analysed from a linguistic and translational perspective, the use of language (i.e. grammatical mistakes, pronunciation errors, unnatural-sounding expressions), was often identified as the main problem in the AI-dubbed clips. Since the target scripts of the clips realised by the 'italiancomedydub' project are not machine-translated but rather created thanks to the collaboration of followers, this result is not to be considered as representative of automatic dubbing as a whole. Nevertheless, this aspect highlights the importance of a good translation for the success of a dubbed product. Therefore, both in the case of fully automated dubbing and in the case of human dubbing, the work of highly skilled translators remains crucial. Whether the target text is to be made from scratch, or only needs post-editing, professionals must be able to go beyond the mere appropriate translation of grammatical structures and syntax. Instead, they should be also able to render paralinguistic features, cultural nuances and ideosyncrasies present in each language.

As Nowell et al. (2017) notes, "in qualitative research, the process of data collection, data analysis, and report writing are not always distinct steps" (2017: 4). This is important to clarify, as it is suggested here, that the findings of the present study are not limited to the answers gathered in the reception study. Prior to the actual interviews, indeed, an informative questionnaire was carried out with the admins of the selected Instagram profiles – chosen as representative examples

of automatic dubbing practices. The answers provided in this stage can be considered as valuable data. In Section 2.3., the projects' workflows, objectives and perspectives were outlined. In particular, the collaborative relationship between the admins of the 'italiancomedydub' project and its followers offers significant insight into the realm of cyberdubbing, which challenges established norms and engage in the creation of content transgressing mainstream production workflows. Prior to the digitalisation and the streaming culture, the role of the viewer was that of a passive consumer of media content. Today, Internet, social networks, ongoing technological advances, and the proliferation of accessible software are reshaping this role. In this sense, studies on reception such as the present one are useful in demonstrating that decisions in the field of AVT must be made with a view to actual "active and socially contextualized" (Biltereyst and Meers 2018: 25) audiences. Indeed, new AVT methods are emerging outside mainstream circuits. Content creators on social networks (and their followers) experiment with new AI systems producing and translating audiovisual material, often using AI-based tools. Together with technological development, the new role of the audience has the potential to heavily impact the entertainment industry.

The answers to the final questions of the interview are crucial for addressing the last research question, i.e. RQ 3 ("What are the opinions and attitudes of native English speakers towards the presence of the new AI-powered tools in the field of audiovisual translation?"). As a matter of fact, the concluding part of the interview was specifically designed to elicit participants' opinions and attitudes toward the potential and challenges associated with AI-driven audiovisual translation. Respondents' comments on this theme revealed a mixed feeling towards the integration of these technologies. Indeed, interviewees proved to be both worried and intrigued by the use of AI in AVT. On one side, remarks about the correct and ethical use of this technology were raised. On the other, a widespread curiosity was registered.

Before concluding, it seems appropriate to list the limitations of the present research. It is important to begin by making an observation concerning the sample used for the study. Indeed, since a limited number of subjects was interviewed, results could be significantly different from those obtained through larger-scale studies. Composition of the sample is another important parameter to be taken into account. In this case, the generational profile of the respondents was very specific: it consisted in highly educated young individuals (23 to 33 years old), all of whom were native (British or American) English speakers. In the case of such small samples, it is important to ensure that the sample is consistent and coherent, so that the research is valid, at least, for that segment of the population. The qualitative research method of the interview is particularly efficient for small samples, as it enables to elicit extended and detailed data otherwise difficult to gather in – for instance – quantitative questionnaires with closed answers. Nevertheless, it should be noted that results obtained here cannot be considered as valid for the formulation of statistical norms. To do so, i.e. confirming the findings of the present study, it seems both appropriate and

interesting to conduct quantitative and larger-scale research in the future. Another limitation that should be mentioned is that results do not concern dubbing in general, as just a language pair (Italian and English) was taken into consideration here. Although some issues were related to the actual technical process of dubbing (e.g. lip sync), overall results could be different when varying the source and target language. A further significant factor to be taken into account is technological development. As evidenced by the answers to the informative questionnaire from the admins of the 'italiancomedydub' page, the (often free) online tools available to users (even those with no computer skills) are subject to sudden changes. What is currently considered as valid and cutting-edge today in the field of AI, may become obsolete in the immediate future. Therefore, this study should only be understood as a snapshot, a description of the current situation.

At the present time, there seem to be very little ongoing research addressing these developments. This contribution, thus, attempted to fill in this gap in the existing literature, through an audience-based reception study. Although it may be early to determine the extent to which AI could potentially contribute to the transformation of AVT in the future (also considering the negative judgements towards the automatically dubbed clips in the interview), it was shown that both audiences and amateur content creators have elevated expectations and curiosity in the subject matter. The implications of this are manifold: with the support of audiences and the benefits that these tools offer (in terms of speed, scalability, and cost-effectiveness), it is possible that professional studios will increasingly turn to automation for the localisation and global diffusion of their content, with consequences in both cultural, and economic terms. Since most of "reception studies have so far focused overwhelmingly on subtitling" (Orrego-Carmona 2019: 378), the findings of this study, despite the aforementioned limitations, have the potential to add new elements of discussion to the discourse on dubbing in AVT studies. In particular, the study may prompt further in-depth and larger-scale future research about English-language dubbing, AI dubbing, and cyberdubbing. As already mentioned, larger-scale investigations still need to be carried out to provide a wider picture of audiences' opinions and attitudes towards the use of AI in AVT. In addition to this, other aspects can be investigated further. One example is the notion of 'voice', very difficult to pin down and especially important in automatic dubbing, both for issues related to copyright in speech synthesis (with the voice cloning technology), and for an artistic and qualitative point of view, given that "voices can play a central part in the experience of watching a film" and "may even determine whether a certain film is perceived as aesthetic and enjoyable by the audience" (Naranjo 2021: 581). Another future direction could entail the experimentation of similar reception studies with different variables in the sample's composition in terms of, for example, age, level of education, or mother tongue. In addition to this, the triangulation of different research methods could be useful. In this context, researchers could consider the Internet as both an object of study and as a resource. This is because, "as audiences become more responsive and

willing to express their views through various social media platforms, big data might shed more light on the reception of translated content in the future" (Orrego-Carmona 2019: 378). Finally, future contributions may employ interdisciplinary approaches (drawing from AVT reception studies, philosophy, law and studies on AI) to explore the use and the implications of AI in audiovisual translation not only from the perspective of enjoyment and appreciation, but also from a more ethical standpoint, which seemed to be one of the main concerns of the participants to the interview in the present study.

APPENDICES

Appendix A: Informative questionnaire

Titolo del progetto:

English-language cyberdubbing in the Age of artificial intelligence: a reception study

Responsabile del Progetto:

Lorenzo Costabile (Università di Pavia)

Data: **Partecipante:**

1) Qual è il nome del vostro progetto?

2) Qual è lo scopo del vostro progetto?

3) Quante persone sono coinvolte nel progetto? Se più di una, quali sono i ruoli di ognuno?

4) Qual è il livello di istruzione delle persone coinvolte nel progetto? (eventualmente, indicare corso di laurea)

5) Qual è l'occupazione delle persone coinvolte nel progetto? Ci sono professionisti nel campo della traduzione/doppiaggio/informatica/intelligenza artificiale?

6) Come si compone il vostro pubblico? (es. maggioranza italiani, maggioranza inglesi)

7) Ci sono pagine alle quali vi ispirate/vi siete ispirati per la creazione dei vostri contenuti? Se sì, quali?

8) Con quale criterio vengono selezionati i video da doppiare?

9) Come vengono doppiati i video? (con quale/i software?)

10) Il processo di adattamento/traduzione è automatico o richiede l'intervento imano? Se sì, descrivete brevemente l'intervento sul processo.

11) La sincronizzazione dell'audio con il video è automatica o richiede l'intervento umano? Se sì, descrivete brevemente l'intervento sul processo.

12) La sintesi vocale delle versioni doppiate è automatica o richiede l'intervento umano? Se sì, descrivete brevemente l'intervento sul processo.

13) Il tono o l'espressività delle voci doppiate è ricavata in automatico richiede l'intervento umano? Se sì, descrivete brevemente l'intervento sul processo.

14) Quali sono le direzioni future del progetto?

15) Siete consapevoli dei rischi legati all'utilizzo di intelligenze artificiali (ad es. deepfake, perdita di posti di lavoro)?

16) Pensate/sperate che il doppiaggio automatico possa diventare una realtà nel campo dell'intrattenimento?

Appendix B: Interview scheme

Part 1: Demographics and viewing habits

- How old are you?
- What is your education level?
- What is your occupation?
- What is your mother tongue? Do you speak any other language?
- How often do you watch audiovisual content such as films and TV series?
- Do you watch more domestic (English-language) or foreign (e.g. Italian) productions?
- When watching a foreign production, what is your preferred audiovisual translation method (subtitling or dubbing)? Why?

Part 2: Reception questions (for each of the clips)

1) Did the voices in this clip sound natural to you? If not, can you describe when or why the voices felt particularly unnatural?

2) How well did the dubbing sync with the lip movements of the actors? Did you notice any distracting or unrealistic mismatch?

3) Do you think the dubbed voices matched the physical appearance and behaviours of the actors on screen?

4) Do you think that the dubbed version managed to capture the original emotional tone of the scene?

5) Overall, were there any moments when the dubbed speech felt unclear or difficult to understand?

Part 3: Final questions

6) What is your opinion on artificial intelligence?

7) Are you aware that AI tools can be used to translate (dub) artistic products such as films and TV series?

8) Have you ever encountered this type of content? If yes, what is your overall impression?

9) Two out of these four clips were made using AI-based tools. Can you identify them?

10) How do you think knowing in advance that some of these clips were made using AI would have affected your reception?

11) Would you be willing to watch AI-dubbed films in the future?

Appendix C: Informed consent

Title of the project:

English-language cyberdubbing in the Age of artificial intelligence: a reception study

Investigator:

Lorenzo Costabile (University of Pavia)

Purpose of study and procedure:

You are being asked to take part in a research study. The purpose of this study is to investigate English-language dubbing manifestations and its associated reception by English native speakers. First, you will watch some brief clips. The clips consist in scenes from Italian films dubbed into English. Then, you will be asked to answer to some related questions. The interview will take between 20 and 30 minutes to complete.

Voluntary participation:

Your participation in this study is voluntary. If you agree to take part in this study, please fill the information below.

Date: **Participant's signature:**

REFERENCES

Abril, C. H. (2015) 'Multilingualism in Tarantino's Inglorious Basterds. Difficulties and strategies for dubbing and subtitling', in *Estudios Franco-Alemanes*, 7, 37-57.

Androutsopoulos, J. (2010) 'Participatory culture and metalinguistic discourse: Performing and negotiating german dialects on YouTube', in D. Tannen and A. M. Trester (eds) *Discourse 2.0. Language and New Media*, 47-71.

Antonelli, G. (2016) *L'italiano nella società della comunicazione 2.0*, Il Mulino, Bologna.

Appel, M. and Prietzel, F. (2022) 'The detection of political deepfakes', in *Journal of Computer-Mediated Communication*, 27(4).

Austin, T. (2002) *Hollywood hype and audiences: Selling and watching popular film in the 1990s*, Manchester: Manchester University Press.

Baldo, M. (2009) 'Dubbing multilingual films *La terra del ritorno* and the Italian-Canadian immigrant experience', in M. Giorgio Marrano, G. Nadiani and C. Rundle (eds) *inTRAlinea Special Issue: the Translation of Dialects in Multimedia*.

Baldry, A. and Thibault, P. J. (2006) *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*, London and Oakville: Equinox Publishing Ltd.

Baños, R. (2023) 'Key challenges in using automatic dubbing to translate educational YouTube videos', in *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

Baños, R. and Díaz-Cintas, J. (2023) 'Exploring new forms of audiovisual translation in the age of digital media: Cybersubtitling and cyberdubbing', in *The Translator*, 30(1), 129–144.

Benson, P. (2021) *Language learning environments: Spatial perspectives on SLA*, Bristol, Blue Ridge Summit: Multilingual Matters.

Berke, L., Albusays, K., Seita, M. and Huenerfauth, M. (2019) 'Preferred appearance of captions generated by automatic speech recognition for deaf and hard-of-hearing viewers', in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*, New York: Association for Computing Machinery, 1-6.

Berruto, G. (2007) 'Sulla vitalità sociolinguistica del dialetto oggi', in G. Raimondi, L. Revelli (eds), *La dialectologie aujourd'hui (Atti del Convegno Internazionale "Dove va la dialettologia?", Saint-Vincent, Aosta, Cogne, 21-24 settembre 2006)*, Alessandria: Edizioni dell'Orso, 133-148

Berruto, G. (2012) *Sociolinguistica dell'italiano contemporaneo*, Roma: Carocci (rev. ed.).

Biltereyst, D. and Meers, P. (2018) 'Film, cinema and reception studies. Revisiting research on audience's filmic and cinematic experience', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 21-41.

Bosseaux, C. (2019) 'Investigating dubbing: Learning from the past, looking to the future', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 48-63.

Brannon, W., Virkar, T. and Thompson B. (2023) 'Dubbing in practice: A large scale study of human localization with insights for automatic dubbing', in *Transactions of the Association for Computational Linguistics*, 11, 419–435.

Bruti, S. (2019) 'Spoken discourse and conversational interaction in audiovisual translation', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 192-208.

Chaume, F. (2004) 'Synchronization in dubbing: A translational approach', in P. Orero (ed.) *Topics in Audiovisual Translation*, Amsterdam: John Benjamins, 35-52.

Chaume, F. (2007) 'Quality standards in dubbing: A proposal', in *TradTerm*, 13, 71-89.

Chaume, F. (2012) *Audiovisual translation: Dubbing*, Manchester: St. Jerome.

Chaume, F. (2013) 'The turn of audiovisual translation. New audiences and new technologies', in *Translation Spaces*, 2, 105-123.

Chaume, F. (2018) 'An overview of audiovisual translation: Four methodological turns in a mature discipline', in *Journal of Audiovisual Translation*, 1(1), 40-63.

Chen, Q., Tan, M., Qi, Y., Zhou, J., Li, Y. and Wu, Q. (2022) 'V2C: Visual voice cloning', in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21210-21219.

Cornu, J.F. (2014) *Le doublage et le sous-titrage. Histoire et esthétique,* Rennes: Presses universitaires de Rennes.

Crabb, M. and Hanson, V.L. (2016) 'Dynamic subtitles: The user experience', in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX 2015), 03-05 June 2015, Brussels, Belgium*, New York: ACM, 103-112.

Crabb, M., Jones, R., Armstrong, M. and Huges, C.J. (2015) 'Online news videos', in *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility [ASSETS '15], 26-28 October 2015, Lisbon, Portugal*, New York: ACM, 215-222.

Da Silva, A. C. (2017) 'On Jakobson's intersemiotic translations in Asterix comics', in *Comparatismi*, (2), 71-81.

de Higes-Andino, I. (2014) 'The translation of multilingual films: Modes, strategies, constraints and manipulation in the Spanish translations of It's a Free World …', in *Linguistica Antverpiensia, New Series. Themes in Translation Studies*, 13, 211–231.

de los Reyes Lozano, J. And Mejías-Climent L. (2023) Beyond the black mirror effect: The impact of machine translation in the audiovisual translation environment, in *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 20, 1–19.

de Renzo, F. (2008) 'Per un'analisi della situazione sociolinguistica dell'Italia contemporanea. Italiano, dialetti e altre lingue', in *Italica*, 85(1), 44–62.

Delabastita, D. (1990) 'Translation and the mass media', in S. Bassnett and A. Lefevere (eds) *Translation, History and Culture*, London: Pinter, 97-109.

Desilla, L. (2014) 'Reading between the lines, seeing beyond the images: An empirical study on the comprehension of implicit film dialogue meaning across cultures', in *The Translator*, 20(2), 194–214.

Di Giovanni, E. (2022) 'Audiovisual translation, audiences and reception', in E. Bielsa (ed.) *The Routledge Handbook of Translation and Media*, 400-411.

Díaz-Cintas, J. (1999) 'Dubbing or subtitling: The eternal dilemma', in *Perspectives*, 7(1), 31-40.

Díaz-Cintas, J. (2012) 'Clearing the smoke to see the screen: Ideological manipulation in audiovisual translation', in *Meta*, 57(2), 279–293.

Díaz-Cintas, J. (2019) 'Audiovisual translation', in E. Angelone, M. Ehrensberger-Dow and G. Massey (eds) *The Bloomsbury Companion to Language Industry Studies*, London: Bloomsbury, 209-230.

Díaz-Cintas, J. and Muñoz Sánchez, P. (2006) 'Fansubs: Audiovisual translation in an amateur environment', in *JoSTrans: The Journal of Specialised Translation*, 6, 37-52.

Díaz-Cintas, J. and Orero, P. (2006) 'Voice-over', in K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), 477-479.

Díaz-Cintas, J. and Remael A. (2007) *Audiovisual translation: Subtitling*, London and New York: Routledge.

Dwyer, T. (2019) 'Audiovisual translation and fandom', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 436-452.

Effendi, J., Virkar, Y., Barra-Chicote, R. and Federico, M. (2022) 'Duration modeling of neural TTS for automatic dubbing', in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8037-8041.

Federico, M., Enyedi, R., Barra-Chicote, R., Giri, R., Isik, U., Krishnaswamy, A. and Sawaf, H. (2020) 'From speech-to-speech translation to automatic dubbing', in *Proceedings of the 17th International Conference on Spoken Language Translation*, 257–264.

Ferguson, G. (2015) 'Introduction: attitudes to English', in A. Linn, N. Bermel and G. Ferguson (eds) *Attitudes towards English in Europe*, Berlin and Boston: Mouton De Gruyter, 3-24.

Floridi, L. (2014) *The fourth revolution: How the infosphere is reshaping human reality*, Oxford: Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018) 'AI4People - An

ethical framework for a good ai society: Opportunities, risks, principles, and recommendations', in *Minds & Machines*, 28, 689–707.

Formentelli, M. and Ghia, E. (2021) '"What the hell's going on?" A diachronic perspective on intensifying expletives in original and dubbed film dialogue', in *Textus, English Studies in Italy*, 1/2021, 47-73.

Franco, E., Matamala, A. and Orero, P. (2010) *Voice-over translation: An overview*, Bern: Peter Lang.

Gambier, Y. (2018) 'Translation studies, audiovisual translation and reception', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 43-66.

Gambier, Y. (2023) 'Audiovisual translation and multimodality: What future?', in *Media and Intercultural Communication: A Multidisciplinary Journal*, 1(1), 1-16.

Genelza, G. G. (2024) 'A systematic literature review on AI voice cloning generator: Game-changer or threat?', in *Journal of Emerging Technologies*, 4(2), 54-61.

Georgakopoulou, P. (2019) 'Technologization of audiovisual translation', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 515-539.

Georgakopoulou, P. (2019) 'Template files: The Holy Grail of subtitling', in *Journal of Audiovisual Translation*, 2(2), 137-160.

Ghia, E. and Pavesi, M. (2021) 'Choosing between dubbing and subtitling in a changing landscape', in *Lingue e Linguaggi*, 46, 161-177.

González Ruiz, V. M. and Cruz García, L. (2021) 'Other voices, other rooms? The relevance of dubbing in the reception of audiovisual products', in *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 6, 219-233.

Guerberof-Arenas, A. (2019) 'Pre-editing and post-editing', in E. Angelone, M. Ehrensberger-Dow and G. Massey (eds) *The Bloomsbury Companion to Language Industry Studies*, London: Bloomsbury, 333-360.

Guerberof-Arenas, A., Moorkens, J. and Orrego-Carmona, D. (2024) 'A Spanish version of EastEnders": A reception study of a telenovela subtitled using MT', in *The Journal of Specialised Translation*, (41), 230–254.

Guillot, M.N. (2019) 'Subtitling on the cusp of its futures', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 31-47.

Guillot, M.N. (2020) 'The pragmatics of audiovisual translation: Voices from within in film subtitling' in *Journal of Pragmatics*, 170, 317-330.

Hall, S. (1973) *Encoding and decoding in the television discourse*, Birmingham: Centre for Contemporary Cultural Studies.

Hayes, L. (2021) 'Netflix disrupting dubbing: English dubs and British accents', in *Journal of Audiovisual Translation*, 4(1), 1–26.

Hayes, L. (2023) 'English dubs: why are anglophone viewers receptive to English dubbing on streaming platforms and to foreign-accent strategies?', in *Íkala, Revista de Lenguaje y Cultura*, 28(2), 1-20.

Heiss, C. (2004). Dubbing multilingual films: a new challenge?, in *Meta*, 49(1), 208–220.

Herbst, T. (1994) *Linguistische aspekte der synchronisation von fernsehserien. phonetik, textlinguistik, übersetzungstheorie*, Tübingen, Niemeyer.

Herbst, T. (1997) 'Dubbing and the dubbed text – Style and cohesion: Textual characteristics of a special form of translation', in A. Trosborg (ed.) *Text Typology and Translation*, Amsterdam and Philadelphia: John Benjamins, 291-308.

Hill, A. (2018) 'Media audiences and reception studies', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 3-19.

Hu, C., Tian, Q., Li, T., Wang, Y., Wang, Y., and Zhao, H. (2021) 'Neural dubber: Dubbing for silent videos according to scripts', in *Neural Information Processing Systems*.

Incalcaterra McLoughlin, L. (2019) 'Audiovisual translation in language teaching and learning', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 483-497.

Incalcaterra McLoughlin, L. and Lertola J. (2011) 'Learn through subtitling: subtitling as an aid to language learning', in L. Incalcaterra McLoughlin, M. Biscio and M.A. Ní Mhainnín (eds) *Audiovisual Translation. Subtitles and Subtitling: Theory and Practicea*, Bern: Peter Lang, 243-263.

Jäckel, A. (2001) 'The subtitling of La Haine: A case study', in Y. Gambier and H. Gottlieb (eds) *(Multi) Media Translation: Concepts, Practices and Research*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 223-235.

Jakobson, R. (1959) 'On linguistic aspects of translation', in R. Brower (ed.) *On Translation*. Cambridge, MA and London: Harvard University Press, 232-239.

Jezek, E. and Sprugnoli, R. (2023) *Linguistica computazionale. Introduzione all'analisi automatica dei testi*, Bologna: Il Mulino.

Kaspsaskis, D. (2020) 'Subtitling, interlingual', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (3rd ed.), London: Routledge, 554-560.

Kim, H., Elgharib, M., Zollhöfer, M., Seidel, H.P., Beeler, T., Richardt, C. and Theobalt, C. (2019) 'Neural style-preserving visual dubbing', in *ACM Transactions on Graphics*, 38(6), 1-13.

Koolstra C. M., Peeters, A. L., and Spinhof, H. (2002) 'The pros and cons of dubbing and subtitling', in *European Journal of Communication*, 17(3), 325-354

Kress, G. (2009) *Multimodality. A social semiotic approach to contemporary communication*, London and New York: Routledge.

Kress, G. and Van Leeuwen, T. (2001) *Multimodal discourse. The modes and media of contemporary communication*, London: Arnold Publishers.

Kruger, J. L. (2020) 'Audio description', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (3rd ed.), London: Routledge, 27-30.

Kruger, J. L. and Doherty, S. (2018) 'Triangulation of online and offline measures of processing and reception in AVT', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 91-177.

Lakew, S.M., Federico, M., Wang, Y., Hoang, C., Virkar, Y., Barra-Chicote, R. and Enyedi R. (2021) 'Machine translation verbosity control for automatic dubbing', in *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*, 7538-7542.

Li, D. (2019) 'Ethnographic research in audiovisual translation', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 383-397.

Lippi-Green, R. (2012) *English with an accent: Language, ideology and discrimination in the United States*, London: Routledge.

Liu, Y., Wei, L., Qian, X., Zhang, T., Chen, S. and Yin, X. (2024) 'M3TTS: Multi-modal text-to-speech of multi-scale style control for dubbing', in *Pattern Recognition Letters*, 179, 158-164.

Mangiron, C. (2018) 'Reception studies in game localisation. Taking stock', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 277-296.

Marleau, L. (1982) 'Les sous-titres… un mal nécessaire', in *Meta*, 27(3), 271-285

Massida, S. and Casarini, A. (2017) 'Sub me do: The development of fansubbing in traditional dubbing countries – The case of Italy', in D. Orrego-Carmona and Y. Lee (eds) Non-Professional Subtitling, Newcastle: Cambridge Scholars Publishing, 63-83.

Matamala, A. (2019) 'Voice-over. Practice, research and future prospects', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 64-81.

Matamala, A., Perego E. and Bottiroli, S. (2017) 'Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain', in *Babel*, 63(3), 423-44.

Matrisciano, S. (2021) 'Il dialetto come marcatore di un nuovo stile imprenditoriale italiano negli economi dello street food', in G. Bernini, F. Guerini and G. Iannàccaro (eds) *La Presenza dei Dialetti Italo-Romanzi nel Paesaggio Linguistico. Ricerche e riflessioni*, Bergamo: Sestante Edizioni, 217-236.

McDonald, P. (2009) 'Miramax, Life is Beautiful, and the Indiewoodization of the foreign-language film market in the USA', in *New Review of Film and Television Studies*, 7(4), 353-375.

McDonnell, E.J., Eagle, T., Sinlapanuntakul, P., Moon, S.H., Ringland, K.E., Froehlich, J.E. and Findlater, L. (2024) '"Caption it in an accessible way that is also enjoyable": Characterizing user-driven captioning practices on TikTok', in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, 492, New York: Association for Computing Machinery, 1–16.

Minutella, V. (2021) *(Re)creating language identities in animated films. dubbing linguistic variation*, Cham: Palgrave Macmillan Cham.

Naranjo, B. (2021) 'The role of emotions in the perception of natural vs. play-acted dubbing: An approach to angry and sad vocal performances', in *Meta*, 66(2), 580-600.

Neves, J. (2019) 'Subtitling for deaf and hard of hearing audiences', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 82-95.

Nida, E. A. and Taber, C. R. (1969) *The theory and practice of translation*, Leiden: E. J. Brill.

Nowell, L. S., Norris, J. M., White, D. E. and Moules, N. J. (2017) 'Thematic analysis: *S*triving to meet the trustworthiness criteria', in *International Journal of Qualitative Methods*, 16, 1-13.

O'Connell, E. (2003) 'What dubbers of children's television programmes can learn from translators of children's books?', in *Meta*, 48 (1-2), 222-232.

O'Hagan, M. (2009) 'Evolution of user-generated translation: *F*ansubs, translation hacking and crowdsourcing', in *The Journal of Internationalization and Localization*, 1(1), 94-121.

O'Hagan, M. (2020) 'Technology, audiovisual translation', in M. Baker and G. Saldanha (eds) *Routledge encyclopedia of translation studies* (3rd ed.), London: Routledge, 565-569.

O'Sullivan C. and Cornu J.F. (2019) 'History of audiovisual translation', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 15-30.

Orrego-Carmona, D. (2016) 'A reception study on non-professional subtitling: Do audiences notice any difference?', in *Across Language and Cultures*, 17(2), 163-181.

Orrego-Carmona, D. (2018) 'New audiences, international distribution, and translation', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 321-342.

Orrego-Carmona, D. (2019) 'Audiovisual translation and audience reception', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 367-382.

Palermo, M. (2023) 'La rappresentazione multimodale dei dialetti su Tik Tok', in *Italiano LinguaDue*, 14(2), 131-139.

Pavesi, M. (2005) *La traduzione filmica. Aspetti del parlato doppiato dall'inglese all'italiano*, Roma: Carocci editore.

Pavesi, M. (2018) 'Translational routines in dubbing: taking stock and moving forward', in I. Ranzato and S. Zanotti (eds) *Linguistic and Cultural Representation in Audiovisual Translation*, London and New York: Routledge, 11-30.

Pavesi, M. (2019) 'Corpus-based audiovisual translation studies: Ample room for development', in in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 315-333.

Pavesi, M. (2020) 'Dubbing', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (3rd ed.), London: Routledge, 156-161.

Pavesi, M. and Ghia, E. (2020) *Informal contact with english. A case study of Italian postgraduate students*, Pisa: Edizioni ETS.

Pavesi, M. and Zamora, P. (2022) 'The reception of swearing in film dubbing: a cross-cultural case study', in *Perpectives*, 30(3), 382–398.

Pederson, J. (2007) 'How is culture rendered in subtitles?', in *MuTra 2005 – Challenges of Multidimensional Translation: Conference Proceedings, Saarbrücken 2-6 May 2005*, 1-18.

Perego, E. (2019) 'Audio description: Evolving recommendations for usable, effective and enjoyable practices', in L. Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation*, London and New York: Routledge, 114-129.

Perego, E., Del Missier, F. and Bottiroli S. (2015) 'Dubbing versus subtitling in young and older adults: Cognitive and evaluative aspects', in *Perspectives*, 23, 1-21.

Perego, E., Del Missier, F. and Stragà, M. (2018) 'Dubbing vs. subtitling: Complexity matters', in *Target*, 30(1), 137-157.

Pérez-González, L. (2014) *Audiovisual translation. Theories, methods and issues* (1st ed.), London: Routledge.

Pérez-González, L. (2007) 'Fansubbing anime: insights into the 'butterfly effect' of globalisation on audiovisual translation', in *Perspectives*, 14(4), 260-277.

Pérez-González, L. (2009) 'Audiovisual translation', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (2nd ed.), London and New York: Routledge, 13-20 .

Pérez-González, L. (2020) 'Audiovisual translation', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (3rd ed.), London and New York: Routledge, 30-34.

Quargnolo, M. (2000) 'Il doppiato italiano', in C. Taylor (ed.) *Tradurre il Cinema: Atti del Convegno organizzato da G. Soria e C. Taylor, 29-30 novembre 1996*, Trieste: Dipartimento di Scienze del Linguaggio, dell'Interpretazione e della Traduzione, 19-21.

Ren, J., Xu, H., He, P., Cui, Y., Zeng, S., Zhang, J., Wen, H., Ding, J., Huang, P., Lyu, L., Liu, H., Chang, Y. and Tang, J. (2024) 'Copyright protection in generative AI: A technical perspective', *arXiv:2402.02333v2*.

Romano, M. (2015) 'Lingua e dialetto nel cinema comico contemporaneo: Checco Zalone e Ficarra e Picone', in G. Marcato (ed.) *Atti del Convegno di Sappada/Plodn (Belluno) 25-30 giugno 2014*, Padova: CLEUP, 1-7.

Romero-Fresco, P. (2009) 'Naturalness in the Spanish dubbing language: A case of not-so-close Friends', in *Meta*, 54(1), 49–72.

Romero-Fresco, P. (2011) *Subtitling through speech recognition: Respeaking*, London: Routledge.

Romero-Fresco, P. (2013) 'Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmaking', in *The Journal of Specialised Translation*, 20, 201-223

Romero-Fresco, P. (2020) 'Subtitling for the deaf and hard of hearing', in M. Baker and G. Saldanha (eds) *Routledge Encyclopedia of Translation Studies* (3rd ed.), London: Routledge, 549-554.

Romero-Fresco, P. and Chaume, F. (2022) 'Creativity in audiovisual translation and media accessibility', in *The Journal of Specialised Translation*, 38, 75-101.

Romero-Fresco, P., and Fryer, L. (2013) 'Could audio-described films benefit from audio introductions? An audience response study', in *Journal of Visual Impairment & Blindness*, 107(4), 287-295.

Rothe, S., Tran, K. and Hussmann, H. (2018) 'Positioning of subtitles in cinematic virtual reality', in G. Bruder, S. Cobb and S. Yoshimoto (eds) *ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, 1-8.

Saboo, A. and Baumann, T. (2019) 'Integration of dubbing constraints into machine translation', in *Proceedings of the Fourth Conference on Machine Translation*, 1, 94–101.

Sahipjohn, N., Gudmalwar, A., Shah, N., Wasnik, P. and Shah, R. (2024) 'DubWise: Video-guided speech duration control in multimodal LLM-based text-to-speech for dubbing', *arXiv:2406.08802*.

Saldaña, J. (2013) *The coding manual for qualitative researchers*, 2nd ed, London: Sage.

Sánchez-Mompeán, S. (2015) 'Dubbing animation into Spanish: Behind the voices of animated characters', in *The Journal of Specialised Translation*, 23, 270-291.

Sánchez-Mompeán, S. (2023) 'Engaging English Audiences in the dubbing experience: A matter of quality or habituation?', in *Íkala, Revista de Lenguaje y Cultura*, 28(2), 1-18.

Schwab, K. (2016) *The fourth industrial revolution*, Cologny: World Economic Forum.

Spiteri Miggiani, G. (2021) 'Exploring applied strategies for english-language dubbing', in *Journal of Audiovisual Translation*, 4(1), 137–156.

Spiteri Miggiani, G. (2024) 'Quality assessment tools for studio and AI-generated dubs and voice-overs', in *Parallèles*, 36(2), 50–70.

Stöckl, H. (2004) 'In between modes: Language and image in printed media', in E. Ventola, C. Charles and M. Kaltenbacher (eds) *Perspectives on Multi-modality* (Document Design Companion Series 6), Amsterdam: Benjamins, 9-30.

Taronna, A. (2016) *Black Englishes. Pratiche linguistiche transfrontaliere Italia-USA*, Verona: Ombre Corte.

Taylor, C. (2012) 'Multimodal texts', in E. Perego (ed.) *Eye Tracking in Audiovisual Translation*, Rome: Aracne Editrice, 13-35.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. and Nießner, M. (2016) 'Face2Face: Real-time face capture and reenactment of RGB videos', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2387-2395.

Thompson, B., Dhaliwal, M., Frisch, P., Domhan, T. and Federico, M. (2024) 'A shocking amount of the web is machine translated: Insights from multi-way parallelism', in *Findings of the Association for Computational Linguistics ACL 2024*, 1763-1775.

Toury, G. (1995) *Descriptive translation studies and beyond*, Amsterdam and Philadelphia: John Benjamins Publishing Company.

Venuti, L. (1995) *The translator's invisibility. A history of translation*, London: Routledge.

Wang, Y. (2023) 'Artificial intelligence technologies in college english translation teaching', in *Journal of Psycholinguistic Research*, 52, 1525–1544.

Wissmath, B., Weibel, D. and Groner, R. (2009) 'Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment', in *Journal of Media Psychology*, 21(3), 114–125.

Wu, Y., Guo, J., Tan, X., Zhang, C., Li, B., Song, R., He, L., Zhao, S., Menezes, A. and Bian, J. (2023) 'VideoDubber: Machine translation with Speech-Aware Length Control for Video Dubbing', in *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11), 13772-13779.

Yang, Y., Shillingford, B., Assael, Y., Wang, M., Liu, W., Chen, Y, Zhang, Y., Sezener, E., Cobo ,L. C., Denil, M., Aytar, Y. and de Freitas, N. (2020) 'Large-scale multilingual audio visual dubbing', *ArXiv, abs/2011.03530*.

Zabalbeascoa, P., Izard, N. and Santamaria L. (2001) 'Disentangling audiovisual translation into Catalan from the Spanish media mesh', in Y. Gambier and H. Gottlieb (eds) *(Multi) Media Translation: Concepts, Practices and Research*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 101-112.

Zanotti, S. (2018) 'Historical approaches to AVT reception. Methods, issues and perspectives', in E. Di Giovanni and Y. Gambier (eds) *Reception Studies and Audiovisual Translation*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 134-156.

Zhang, Z., Yang, Q., Wang, D., Huang, P., Cao, Y., Ye, K. and Hao, J. (2024) 'Mitigating unauthorized speech synthesis for voice protection', *arXiv:2410.20742*.

SITOGRAPHY

AI Act, *Home page*, European Commission, available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed: 17/09/24).

Amoeba Sisters en Español, *Home page*, YouTube, available at: https://www.youtube.com/@AmoebaSistersEspanol (accessed: 16/09/24).

Amoeba Sisters, *Home page*, YouTube, available at: https://www.youtube.com/@AmoebaSisters (accessed: 16/09/24).

AVTE (2024) *AVTE statement on generative artificial intelligence*, April 29, available at: https://avteurope.eu/2024/04/29/avte-statement-on-generative-ai/ (accessed: 15/10/24).

Box Office Mojo by IMDbPro, *Genre keyword: foreign language*, available at: https://www.boxofficemojo.com/genre/sg4208980225/ (accessed 21/10/24).

Bryant, M. (2021) *Where have all the translators gone?*, The Guardian, November 14, available at: https://www.theguardian.com/tv-and-radio/2021/nov/14/where-have-all-the-translators-gone (accessed: 17/09/24).

Deedpdub, *Home page,* available at: https://deepdub.ai/ (accessed: 16/09/24).

Direzione generale cinema e audiovisivo (2023) *Pubblicato il report "Tutti I numeri del cinema italiano – anno 2021*, April 21, Ministero della Cultura, available at: https://cinema.cultura.gov.it/notizie/pubblicato-il-report-tutti-i-numeri-del-cinema-italiano-anno-2021/#:~:text=La%20Direzione%20generale%20Cinema%20e,%2Dpandemia%20(325%20nel%202019) (accessed 20/10/24).

Dubbing Wiki, *Home*, Fandom, available at: https://dubbing.fandom.com/wiki/Dubbing_Wikia (accessed: 20/10/24).

Dubbing Wiki, *Italian films*, Fandom, available at: https://dubbing.fandom.com/wiki/Category:Italian_Films (accessed: 20/10/24).

Dubbing Wiki, *Rose Island*, Fandom, available at: https://dubbing.fandom.com/wiki/Rose_Island (accessed 21/10/24).

Fandom, *What is Fandom?*, [available at: https://about.fandom.com/what-is-fandom (accessed: 20/10/24).

Fuster, J. (2024) *For voice actors, the race against AI has already begun*, The Wrap, March 4, available at: https://www.yahoo.com/entertainment/voice-actors-race-against-ai-140000855.html (accessed: 17/09/24).

Gastaldi, S. (2021) *"Strappare lungo i bordi": mescolanza di generi e tanta filosofia. Ma il doppiaggio in inglese lascia a desiderare*, Linkiesta, November 21, available at:

https://www.linkiesta.it/blog/2021/11/strappare-lungo-i-bordi-mescolanza-di-generi-e-tanta-filosofia-ma-il-doppiaggio-in-inglese-lascia-a-desiderare/ (accessed: 06/09/24).

HiRezTV (2024) *Donald Trump – Many Men (50 Cent Remix)*, YouTube, available at: https://www.youtube.com/watch?v=f90BL4uVIag&ab_channel=HiRezTV (accessed: 17/09/24).

italiancomedydub, Instagram, available at: https://www.instagram.com/italiancomedydub/ (accessed: 17/09/24).

Kaufman, A. (1999) *Editorial: Life Isn't Beautiful anymore, it's dubbed*, IndieWire, August 23, available at: https://www.indiewire.com/news/general-news/editorial-life-isnt-beautiful-anymore-its-dubbed-82123/ (accessed 21/10/24).

Kelly, L. (2023) *Why Gen Z is watching TV with the subtitles on*, The Times, October 29, available at: https://www.thetimes.com/culture/tv-radio/article/why-gen-z-is-watching-tv-with-the-subtitles-on-gv0fws395 (accessed: 26/07/24).

Koehler, R. (1999) *Life Is Beautiful (English dubbed version)*, Variety, August 27, available at: https://variety.com/1999/film/reviews/life-is-beautiful-english-dubbed-version-1117752050/ (accessed 21/10/24).

Lee, W. (2022) *Why dubbing has become more crucial to Netflix's business*, Los Angeles Times, February 28, available at: https://www.latimes.com/entertainment-arts/business/story/2022-02-28/why-dubbing-has-become-more-crucial-to-netflixs-business (accessed 06/09/24).

Mackenzie J. and Choi L. (2024) Inside the deepfake porn crisis engulfing Korean schools, BBC, September 3, available at: https://www.bbc.com/news/articles/cpdlpj9zn9go (accessed: 17/09/24).

Molinari, P. (2024) *Europee: Decaro a lezione di dialetti, lo spot spopola sui social*, AGI, May 25, available at: https://www.agi.it/politica/news/2024-05-25/elezioni-europee-antonio-decaro-impara-dialetti-26521547/ (accessed: 01/10/24).

Movieplayer.it, *Film con maggiori incassi in Italia*, available at: https://movieplayer.it/film/boxoffice/italia/di-sempre/ (accessed: 01/10/24).

Mymovies.it, *Premi e nomination La vita è bella*, available at: https://www.mymovies.it/film/1997/lavitaebella/premi/ (accessed 21/10/24).

O'Connor J.F. and Moxley E. (2023) *Our approach to responsible AI innovation*, YouTube Official Blog, November 14, available at: https://blog.youtube/inside-youtube/our-approach-to-responsible-ai-innovation/ (accessed: 16/09/24).

Pogue, D. (2024) *Read all about it: The popularity of turning captions on*, CBS News, March 10, available at: https://www.cbsnews.com/news/subtitles-why-most-people-turn-tv-captions-on/ (accessed: 26/07/24).

Raffaelli, A. (2010) 'Fascismo, lingua del', in *Enciclopedia dell'italiano*, Treccani, available at: https://www.treccani.it/enciclopedia/lingua-del-fascismo_(Enciclopedia-dell'Italiano)/ (accessed: 01/10/24).

Roxborough, S. (2019) *Netflix's global reach sparks dubbing revolution: "the public demands it*, The Hollywood Reporter, August 13, available at: https://www.hollywoodreporter.com/tv/tv-news/netflix-s-global-reach-sparks-dubbing-revolution-public-demands-it-1229761/ (accessed: 06/09/24).

sandrumasi, Instagram, available at: https://www.instagram.com/sandrumasi/ (accessed: 17/09/24).

Satariano A. and Mozur P. (2023) *The people onscreen are fake. the disinformation is real*, The New York Times, February 7, available at: https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html (accessed 17/09/24).

Scherer, M. (2024) *The SAG-AFTRA strike is over, but the AI fight in hollywood is just beginning*, Center for democracy & technology, January 4, available at: https://cdt.org/insights/the-sag-aftra-strike-is-over-but-the-ai-fight-in-hollywood-is-just-beginning/ (accessed: 17/09/24).

Sky News en Español, *Home page*, YouTube, available at: https://www.youtube.com/@skynewsespanol (accessed: 16/09/24).

speechif.ai, Instagram, available at: https://www.instagram.com/speechif.ai/ (accessed: 17/09/24).

Team Papercup (2024) *Will AI replace voice actors?*, Papercup, July 10, available at: https://www.papercup.com/blog/will-ai-replace-voice-actors (accessed: 16/09/24).

The White House (2023) *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*, October 30, available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (accessed: 15/10/24).

Thompson, S.A. (2024) *How 'deepfake Elon Musk' became the Internet's biggest scammer*, The New York Times, August 14, available at: https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html (accessed: 17/09/24).

Time (2021) 'A fix for film dubbing. Flawless AI TrueSync', in *The best inventions of 2021*, Time, available at: https://time.com/collection/best-inventions-2021/6112554/flawless-ai-truesync/ (accessed: 17/09/24).

Todasco, M. (2023) *Deep fakes for all: the proliferation of AI voice cloning*, Medium, August 18, available at: https://medium.com/@todasco/deep-fakes-for-all-the-proliferation-of-ai-voice-cloning-ecee0a461dac (accessed: 07/10/24).

u/MorganAndMerlin (2023) 'English Dubs on tv shows/movies originally in another language are appallingly bad on Netflix. Is it me? Did I not realize until now just because of sheer volume of media content?', in *r/netflix*, Reddit, September 9, available at: https://www.reddit.com/r/netflix/comments/16efetg/english_dubs_on_tv_showsmovies_originally_in/ (accessed: 06/09/24).

Welk B. and Maglio T. (2024) *This new AI tool could have made Joaquin Phoenix's 'Napoleon' sound French*, IndieWire, March 13, available at: https://www.indiewire.com/news/business/ai-tool-joaquin-phoenix-french-napoleon-deepdub-accent-control-1234958496/ (accessed: 17/09/24).

FILMOGRAPHY

A bigger splash (2015) Luca Guadagnino, IMDb entry:
https://www.imdb.com/title/tt2056771/?ref_=ext_shr_lnk

Black mirror (2011-) [TV series] Charlie Brooker, IMDb entry:
https://www.imdb.com/title/tt2085059/?ref_=ext_shr_lnk

Basic instinct (1992) Paul Verhoeven, IMDb entry:
https://www.imdb.com/title/tt0103772/?ref_=ext_shr_lnk

Bram Stoker's Dracula (1992) Francis Ford Coppola, IMDb entry:
https://www.imdb.com/title/tt0103874/?ref_=ext_shr_lnk

Chicken run (2000) Peter Lord and Nick Park, IMDb entry:
https://www.imdb.com/title/tt0120630/?ref_=ext_shr_lnk

Crouching tiger, hidden dragon [Simplified Chinese: 卧虎藏龙; pinyin: Wòhǔ Cánglóng] (2000) Ang Lee, IMDb entry:
https://www.imdb.com/title/tt0190332/?ref_=ext_shr_lnk

Do the right thing (1989) Spike Lee, IMDb entry:
https://www.imdb.com/title/tt0097216/?ref_=ext_shr_lnk

Don Juan (1926) Alan Crosland, IMDb entry:
https://www.imdb.com/title/tt0016804/?ref_=ext_shr_lnk

Dr. Strangelove or: how i learned to stop worrying and love the bomb (1964) Stanley Kubrick, IMDb entry:
https://www.imdb.com/title/tt0057012/?ref_=ext_shr_lnk

Ennio (2021) Giuseppe Tornatore, IMDb entry:
https://www.imdb.com/title/tt3031654/

Fantastic Mr. Fox (2009) Wes Anderson, IMDb entry:
https://www.imdb.com/title/tt0432283/?ref_=ext_shr_lnk

Goodfellas (1990) Martin Scorsese, IMDb entry:
https://www.imdb.com/title/tt0099685/?ref_=ext_shr_lnk

Indagine su un cittadino al di sopra di ogni sospetto (1970) Elio Petri, IMDb entry:
https://www.imdb.com/title/tt0065889/?ref_=ext_shr_lnk

Inglorious basterds (2009) Quentin Tarantino, IMDb entry:
https://www.imdb.com/title/tt0361748/?ref_=ext_shr_lnk

Isle of dogs (2018) Wes Anderson, IMDb entry:
https://www.imdb.com/title/tt5104604/?ref_=ext_shr_lnk

Kung Fu Panda (2008) Mark Osborne and John Stevenson, IMDb entry:
https://www.imdb.com/title/tt0441773/?ref_=ext_shr_lnk

La casa de papel (2017-2021) [TV series] Álex Pina, IMDb entry:
https://www.imdb.com/title/tt6468322/?ref_=ext_shr_lnk
La gabianella e il gatto (1998) Enzo D'Alò, IMDb entry:
https://www.imdb.com/title/tt0122735/?ref_=ext_shr_lnk
La haine (1995) Mathieu Kassovitz, IMDb entry:
https://www.imdb.com/title/tt0113247/?ref_=ext_shr_lnk
L'incredibile storia dell'isola delle rose (2020) Sidney Sibilia, IMDb entry:
https://www.imdb.com/title/tt10287954/?ref_=ext_shr_lnk
La vita è bella (1997) Roberto Benigni, IMDb entry;
https://www.imdb.com/title/tt0118799/?ref_=ext_shr_lnk
Lo spietato (2019) Renato De Maria, IMDb entry:
https://www.imdb.com/title/tt9239888/?ref_=ext_shr_lnk
Natural born killers (1994) Oliver Stone, IMDb entry:
https://www.imdb.com/title/tt0110632/?ref_=ext_shr_lnk
Quo vado? (2016) Gennaro Nunziante, IMDb entry:
https://www.imdb.com/title/tt5290524/?ref_=ext_shr_lnk
Shrek 2 (2004) Andrew Adamson, Kelly Asbury and Conrad Vernon, IMDb entry:
https://www.imdb.com/title/tt0298148/?ref_=ext_shr_lnk
Squid game [Korean: *오징어 게임*; Romanization: *Ojing-eo Geim*] (2021-) [TV series] Hwang Dong-hyuk, IMDb entry:
https://www.imdb.com/title/tt10919420/?ref_=ext_shr_lnk
Strappare lungo i bordi (2021) [TV series] Zerocalcare, IMDb entry:
https://www.imdb.com/title/tt15614372/?ref_=ext_shr_lnk
The big bang theory (2007-2019) [TV series] Chuck Lorre and Bill Prady, IMDb entry:
https://www.imdb.com/title/tt0898266/?ref_=ext_shr_lnk
The gentlemen (2019) Guy Ritchie, IMDb entry:
https://www.imdb.com/title/tt8367814/?ref_=ext_shr_lnk
The jazz singer (1927) Alan Crosland, IMDb entry:
https://www.imdb.com/title/tt0018037/?ref_=ext_shr_lnk
The killing (1956) Stanley Kubrick, IMDb entry:
https://www.imdb.com/title/tt0049406/?ref_=ext_shr_lnk
The Polar Express (2004) Robert Zemeckis, IMDb entry:
https://www.imdb.com/title/tt0338348/?ref_=ext_shr_lnk
The room next door (2024) by Pedro Almodóvar, IMDb entry:
https://www.imdb.com/title/tt29439114/?ref_=ext_shr_lnk
The Simpsons (1989-) [TV series] James L. Brooks, Matt Groening and Sam Simon, IMDb entry:

https://www.imdb.com/title/tt0096697/?ref_=ext_shr_lnk

Totò, Peppino e la… malafemmina (1956) Camillo Mastrocinque, IMDb entry:

https://www.imdb.com/title/tt0049866/?ref_=ext_shr_lnk

Toy story (1995) John Lasseter, IMDb entry:

https://www.imdb.com/title/tt0114709/?ref_=ext_shr_lnk

Tre uomini e una gamba (1997) Aldo Baglio, Giacomo Poretti and Giovanni Storti, IMDb entry:

https://www.imdb.com/title/tt0135007/?ref_=ext_shr_lnk

Triangle of sadness (2022) Ruben Östlund, IMDb entry:

https://www.imdb.com/title/tt7322224/?ref_=ext_shr_lnk

Vanda (2022-) [TV series] Patrícia Müller, IMDb entry:

https://www.imdb.com/title/tt15316232/?ref_=ext_shr_lnk