



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Università degli Studi di Modena e Reggio Emilia

DIPARTIMENTO DI STUDI LINGUISTICI E CULTURALI

CORSO DI LAUREA MAGISTRALE IN
LINGUE, CULTURE, COMUNICAZIONE

*On the Rendition of Subtitles: A Corpus-Based Analysis
of two Contemporary Comedies*

Prova finale di:

Sara Busco

Relatrice:

Laura Gavioli

Correlatore:

Federico Corradini

Anno Accademico 2025-2026

ABSTRACT

The present thesis investigates the translation of fictional orality in the Italian subtitles of two contemporary English-language comedy series, *The Marvelous Mrs. Maisel* (2017-2023) and *Fleabag* (2016-2019), through a corpus-based analysis of fan-produced subtitles. The choice of the two series is grounded on the perception that they show contrasting linguistic features within the comedy genre: *The Marvelous Mrs. Maisel* seems characterised by lexically dense fast-paced dialogue and Yiddish realia, whereas speech fragmentation and marked language seem to feature *Fleabag*. Together, they constitute a productive framework for the investigation of how distinctive features of comedy dialogues are managed in subtitle translation.

The study builds on theoretical frameworks of Audiovisual Translation Studies and Corpus-Based Translation Studies. Two parallel self-compiled corpora were analysed using Sketch Engine, one corpus for each TV series, including all the episodes in both English and Italian. The research adopts a predominantly quantitative approach and the analytical tools employed include frequency lists, keyword analysis and parallel concordance queries.

Frequency lists reveal a systematic token reduction in the Italian target texts, consistent with the simplification and condensation tendencies documented in AVT research. Keyword analysis identifies distinctive linguistic features of each corpus, providing a starting point for the selection of the dialogic phenomena to examine in detail. These include the systematic omission of discourse markers as a primary mechanism of lexical reduction in both series; profanity and slang in *Fleabag*, revealing an unexpected dominant tendency towards direct equivalents rather than neutralisation; Yiddish realia in *The Marvelous Mrs. Maisel*, for which strategies differ based on the category of realia, with dominant tendencies towards repetition or absolute universalisations. Finally, lexical density and speech fragmentation are examined through parallel concordance and punctuation frequency analyses.

The results of this research demonstrate that neutralisation cannot be considered a universal tendency in subtitle translation, rather it is context-dependent and varies according to the linguistic features and their relevance for characterisation, together with the constraints of the audiovisual format. Moreover, the comparison between fan-produced and professionally translated subtitles reveals that fansubs are generally more inclined towards creative and foreignising solutions, though with notable exceptions in specific contexts.

This study contributes to corpus-based research on comedy subtitling by offering a systematic analysis of two linguistically different series of the same genre, and by identifying orientations

to the rendition of language features such as the treatment of discourse markers, profanity, slang and realia, that constitute productive grounds for future qualitative investigation.

RIASSUNTO

La presente tesi analizza la traduzione dei dialoghi nella resa italiana dei sottotitoli di due serie televisive comiche contemporanee, *The Marvelous Mrs. Maisel* (2017-2023) e *Fleabag* (2016-2019), attraverso un'analisi basata su corpora di sottotitoli realizzati da fan (fansub). La scelta delle due serie si basa sull'osservazione di caratteristiche linguistiche contrastanti all'interno del genere comedy: *The Marvelous Mrs. Maisel* sembra distinguersi per la densità e il ritmo incalzante dei dialoghi e per la presenza di *realia* Yiddish, mentre un discorso frammentato e un linguaggio marcato sembrano caratterizzare *Fleabag*. Insieme, costituiscono un quadro d'analisi produttivo per esaminare come le caratteristiche peculiari dei dialoghi vengano gestite nella traduzione dei sottotitoli.

La ricerca si basa sui quadri teorici degli studi sulla traduzione audiovisiva e sulla traduzione basata sui corpora. Nello specifico, due corpora paralleli autocompilati sono stati analizzati tramite Sketch Engine, un corpus per ciascuna serie, contenente tutti gli episodi sia in inglese che in italiano. Lo studio adotta un approccio prevalentemente quantitativo e gli strumenti analitici impiegati includono liste di frequenza, analisi delle parole chiave e delle concordanze parallele.

Le liste di frequenza rivelano una riduzione sistematica delle occorrenze nei testi di arrivo in italiano, in linea con le tendenze di semplificazione e condensazione documentate nella ricerca sulla traduzione audiovisiva. L'analisi delle parole chiave identifica i tratti linguistici distintivi di ciascun corpus, fornendo un punto di partenza per la selezione dei fenomeni dialogici da esaminare nel dettaglio. Tali fenomeni includono l'omissione sistematica dei marcatori del discorso come meccanismo primario di riduzione lessicale in entrambe le serie; slang e volgarità in *Fleabag*, con un'inaspettata tendenza dominante verso l'uso di equivalenti diretti piuttosto che verso la neutralizzazione; i *realia yiddish* in *The Marvelous Mrs. Maisel*, per i quali le strategie differiscono con tendenze dominanti verso ripetizione o universalizzazione assoluta. Infine, la densità lessicale e la frammentazione del parlato vengono esaminate attraverso analisi delle concordanze parallele e della frequenza della punteggiatura.

I risultati di questa ricerca dimostrano che la neutralizzazione non può essere considerata una tendenza universale nella traduzione dei sottotitoli, bensì è dipendente dal contesto, dalle caratteristiche linguistiche e dalla loro rilevanza per la caratterizzazione dei personaggi, oltre che dai vincoli del formato audiovisivo. Inoltre, il confronto tra i sottotitoli prodotti dai fan e quelli tradotti professionalmente rivela una tendenza dei primi verso soluzioni più creative, sebbene non manchino eccezioni in contesti specifici.

Questo studio contribuisce alla ricerca basata su corpora sulla sottotitolazione nelle comedy, offrendo un'analisi sistematica di due serie linguisticamente differenti, e identificando i principali orientamenti traduttivi per la resa di fenomeni linguistici come marcatori del discorso, linguaggio profano, slang e realia, che costituiscono terreni produttivi per future indagini qualitative.

RÉSUMÉ

Cette thèse analyse la traduction des dialogues dans les sous-titres italiens de deux séries télévisées comiques contemporaines en langue anglaise, *The Marvelous Mrs. Maisel* (2017-2023) et *Fleabag* (2016-2019), à travers une analyse basée sur des corpus de sous-titres produits par des fans (fansubs). Le choix des deux séries repose sur l'observation de caractéristiques linguistiques contrastées dans le genre comique : *The Marvelous Mrs. Maisel* semble se caractériser par des dialogues denses et au rythme soutenu, ainsi que par la présence de *realia yiddish* ; tandis que la fragmentation du discours et le langage marqué semblent caractériser *Fleabag*. Ensemble, elles constituent un cadre d'analyse efficace pour examiner la manière dont les traits distinctifs des dialogues comiques sont gérés dans la traduction des sous-titres.

La recherche repose sur les cadres théoriques des études en traduction audiovisuelle et en linguistique de corpus. Deux corpus parallèles auto-compilés, un pour chaque série et comprenant l'ensemble des épisodes en anglais et en italien, ont été analysés en utilisant Sketch Engine. Cette étude adopte une approche principalement quantitative et les outils analytiques utilisés incluent les listes de fréquence, l'analyse des mots-clés et l'étude des concordances parallèles. Les listes de fréquence montrent une réduction systématique des occurrences dans les textes cibles, conformément aux tendances de simplification et condensation documentées par la recherche en AVT. L'analyse des mots-clés met en évidence les traits linguistiques distinctifs de chaque série, en fournissant un point de départ pour la sélection des phénomènes d'oralité à examiner. Ces phénomènes comprennent l'omission systématique des marqueurs discursifs comme principal mécanisme de réduction lexicale ; l'argot et les vulgarités dans *Fleabag*, avec une tendance surprenante à recourir à des équivalents directs plutôt qu'à la neutralisation ; et les *realia yiddish* dans *The Marvelous Mrs. Maisel*, pour lesquels les stratégies traductives varient, avec une prédominance de la répétition ou de l'universalisation absolue. Enfin, la densité lexicale et la fragmentation du discours sont examinées à travers l'analyse des concordances parallèles et de la fréquence de la ponctuation.

Les résultats de cette recherche montrent que la neutralisation ne peut pas être considérée comme une tendance universelle dans la traduction des sous-titres ; ça dépend plutôt du contexte, des caractéristiques linguistiques et leur importance pour la caractérisation des personnages, ainsi qu'en fonction des limitations du format audiovisuel. De plus, la comparaison entre les sous-titres produits par les fans et ceux réalisés par des professionnels montre une tendance des fansubbers à privilégier des solutions plus créatives, bien que certaines exceptions apparaissent.

Cette étude contribue à la recherche basée sur les corpus dans le domaine du sous-titrage des séries comiques en proposant une analyse systématique de deux séries linguistiquement distinctes appartenant au même genre et en identifiant des orientations quant au traitement de caractéristiques linguistiques telles que les marqueurs discursifs, les grossièretés, l'argot et les *realia*, ce qui ouvre des perspectives particulièrement fécondes pour de futures investigations qualitatives.

TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER 1 - AUDIOVISUAL TRANSLATION AND RENDERING SUBTITLES	5
1.1 Audiovisual Translation Studies: History and Development.....	5
1.2 Audiovisual Translation as a Multimodal Language	6
1.3 Modes of Audiovisual Translation: Revoicing and Captioning.....	8
1.3.1 Revoicing: Dubbing and Voice-over	8
1.3.2 Captioning: Subtitling	9
1.3.3 AVT: Professional vs. Amateur Translators	11
1.4 Subtitling: Technical Constraints and Issues of Fidelity to the ST.....	13
1.4.1 Subtitling: Punctuation Conventions in AVT.....	14
1.5 From Spoken to Written Dialogue: Diamesic Shift and Prefabricated Orality	15
1.5.1 Marked Language in Subtitling	18
1.5.2 Slang and Profanities.....	19
1.6 Translating Culture: Realia and Culture-Specific Items	20
1.7 Rendering Humour: the Issue of Untranslatability	24
1.8 Theoretical Synthesis and Analytical Framework	26
CHAPTER 2 - CORPUS-BASED LINGUISTICS AND TRANSLATION STUDIES: A THEORETICAL FRAMEWORK.....	28
2.1 Translation Studies: Premise	28
2.2 A methodological distinction: Corpus-based vs. Corpus-driven approach.....	29
2.3 Taxonomy and Corpus Design	30
2.4 Corpus Design Issues: Representativeness, size, authenticity, sampling, balance, copyright.....	32
2.5 Corpus Pattern Analysis: Norm vs. Creativity.....	33
2.6 Computational Tools for Linguistic Analysis: Sketch Engine	35
2.6.1 Frequency Lists and Keyword Analysis	35
2.6.2 Concordance, Corpus Query Language	36

2.6.3 Collocations and Word Sketches.....	37
2.7 Using Corpus Analysis in Audiovisual Translation	38
2.8 Corpus Design and Selection Criteria.....	39
2.8.1 Corpus Size	39
2.8.2 Data Sources and Fan-Produced Subtitles	40
2.8.3 Pre-processing, Compiling and Metadata Structure.....	41
2.8.4 Limitations of the Corpus Design	41
2.9 Data, Analytical Tools and Methodological Summary	42
CHAPTER 3 - TWO CASES OF STUDY: AN ANALYSIS OF <i>THE MARVELOUS MRS. MAISEL</i> AND <i>FLEABAG</i>	44
3.1 Starting point: Keyword Analysis.....	44
3.2 Wordlist and token count.....	46
3.2.1 <i>The Marvelous Mrs Maisel</i> : Fast-Talking and Lexical Density	48
3.2.2 Frequency Lists: an Analysis of Punctuation Marks	51
3.2.3 <i>Fleabag</i> : Speech Fragmentation and Interrupted Sentences	52
3.3 Condensation and Neutralisation: a Discourse Markers analysis	54
3.4 <i>Fleabag</i> : the Case of Profanities.....	61
3.5 <i>Fleabag</i> : the Translation of Marked Language.....	64
3.5.1 <i>Fleabag</i> : Rendering Idiomatic British Expressions.....	65
3.5.2 <i>Fleabag</i> : Rendering British Realia.....	69
3.5.3 <i>Fleabag</i> : Rendering Wordplays and Jokes.....	70
3.5.4 <i>Fleabag</i> : Concluding Remarks on Marked Language	73
3.6 <i>The Marvelous Mrs. Maisel</i> : Rendering Yiddish Realia	74
3.6.1 Yiddish Realia: a Qualitative Analysis of Selected Cases	77
3.6.2 Yiddish Realia: Concluding Remarks	83
CONCLUSIONS.....	85
BIBLIOGRAPHY.....	89

ONLINE REFERENCES	93
--------------------------------	----

INTRODUCTION

This thesis grows from the desire to combine two personal interests: Audiovisual Translation, which has always been under my radar, and the practical use of Corpus Linguistics to which I have been introduced during my recent academic years and was curious to explore. After catching a glimpse of how corpora tools can be used as devices by translators in specialised fields, I have asked myself whether this instrument could be productively adopted in subtitling. Studies of the last 40 years have shown that translation has long surpassed its connotation of mechanical word-for-word transfer from one language system to another. Instead, it has become more and more pivotal in today's globalised society as a multifaceted process of intercultural negotiations heavily influenced by social and contextual discourse features.

This ever-changing worldwide context has pressured people to adapt to an extremely multicultural society in which both interlingual communication and the role of translators have become essential. As language represents and evolves together with its respective culture and society, it is inevitably shaped by the ongoing development of digitisation in every field of life. Notably, the diffusion of digital streaming platforms has transformed audiovisual products, such as TV series and films, into one of the most renowned channels for global communication, as well as means of representing one's identity. Indeed, characters in audiovisual products, nowadays, have likewise evolved: their discourse is permeated by socio-cultural markers, non-standard registers and colloquial language, in an attempt to simulate a form of spontaneous, yet fabricated, orality. Furthermore, the field of AVT is becoming more and more "free, multilingual and diverse" (Chaume, 2013, p. 121): alongside the well-established modes of subtitling and dubbing, new revoicing and captioning modes, such as Audio Description and Subtitling for the Deaf and Hard of Hearing, have developed in order to meet the accessibility needs of audiences with sensory impairments.

However, the challenges that render the field of subtitling extremely productive for this examination are not only caused by this attempt to simulate the naturalness of dialogues into written forms. Spatial and temporal technical constraints, in fact, remain intrinsic to the nature of the medium per se and often clash with the reproduction of these new stylistic features. That is why when rendering into a target language translators, and subtitlers more specifically, are constantly faced with a pressure for finding a balance between textual reduction and preservation of the unique voice of the original. This is precisely where the purpose of this analysis stems from: investigating how these language features are systematically rendered into

subtitled lines of text in a target culture and what strategies the translators resort to. This systematic analysis is based on empirical data and applies Corpus Linguistics methodologies to the study of two selected datasets bound by the same TV genre. This study aims at achieving an analysis of contemporary comedy speech through the examination of two TV series: *The Marvelous Mrs. Maisel* (Amazon Prime Video, 2017-2023) and *Fleabag* (BBC Three / Amazon Prime Video, 2016-2019). I chose to focus on comedies as this genre can be particularly challenging for audiovisual translators as the faithfulness of its renditions relies on a series of linguistic features such as prefabricated dialogue, marked language and culture-specific items that resist the technical limitations of the format.

As a matter of fact, these features immediately caught my eye as I was watching the episodes; it was evident that these two selected female-led audiovisual products while sharing general similarities in their humorous nature of speech, also presented distinct linguistic traits suitable for formulating hypotheses to test in this research. On the one hand, *The Marvelous Mrs. Maisel* immediately captured my attention with its rapid and lexically dense speech rooted in stand-up comedy traditions. On the other hand, *Fleabag* presented fragmented conversations and continuous hesitations and interruptions. Both discourses, however, are culturally marked: the former is permeated by Yiddish realia that constitute the ideological background of the characters in the Ashkenazi community; the latter by an intensive use of slang and profanities. Together, these two series constitute a productive framework for my study of the rendition of comedic dialogue and its peculiarities in subtitling.

Methodologically this study is based on frameworks of Audiovisual Translation (Gottlieb, 1992, 1994; Díaz-Cintas and Remael, 2007; Chaume, 2013), Descriptive Translation Studies (Toury, 2012) and Corpus-Based Translation Studies (Baker, 1993; Zanettin, 2012; Hanks, 2012; Arias-Badia, 2015, 2020, 2021; Pavesi, 2019; Laviosa, 2021). By applying corpus linguistics methodologies, this work aims at quantitatively evaluating to what extent these linguistic features are preserved or systematically neutralised in the Italian rendering of fan-produced subtitles. This way, the corpus research confirms or refutes the neutralisation tendency documented in AVT research as a pattern across translation in order to fit into the technical boundaries.

In this regard, a quantitative macro-analysis of each dataset allows isolating proof of these specific linguistic phenomena, as well as evaluating text reduction and condensation hypothesis supported by the format. If these assumptions are detected in the findings of the present research, further exploration should be carried out in order to justify them and identify which elements are more commonly subject to deletion or reformulation strategies and

therefore contribute to such textual diminution. In order to answer these research questions, the present study adopts a primarily corpus-based approach, even though, on some occasions, corpus-driven tools can be employed to examine the pre-identified phenomena. For instance, the keyword tool can be used to automatically isolate words for corpus-driven research, but was adopted here to confirm the occurrence of the pre-observed linguistic phenomena in a corpus-based approach. Two parallel DIY corpora have been compiled, one for each TV series including the fan-produced version of the subtitles of the entire episodic run in both English and Italian. Fansubs, downloaded from OpenSubtitles and pre-processed through Subtitle Edit, are examined through quantitative analytical tools on Sketch Engine, accompanied by selected qualitative investigation of specific instances.

The choice to adopt amateur fan-produced subtitles, while strongly influenced by copyright limitations, inevitably affects the findings of the present study. Given the widespread popularity of both series, a variety of translations has been produced: fansubs, in particular, offer meaningful insight into how non-professional translators perceive these conversational markers and to what extent they consider these features crucial to preserve the meaning and humorous effect of the show. This opens up the present work to occasional comparative analysis with manually reported professional subtitling renditions, as well as alternative personal suggestions.

The present thesis is structured in three main chapters. Chapter 1 introduces the theoretical framework of Audiovisual Translation Studies with a brief overview of its development over time, the two main modalities, with a focus on captioning and its technical constraints, as the core object of the present study. Furthermore, sections will be dedicated to the concepts of diamesic shift, which describes the semiotic passage from spoken code of the actor's delivery to written code of the subtitle lines on screen, and the issues of pre-fabricated dialogue, constituted by marked language and culture-specific items. As the work chooses to deal with humorous texts, a brief section will also be dedicated to how these marked features influence the comic effect of the texts and the assumption of untranslatability of humour.

Chapter 2 presents the methodological framework of corpus-based approach starting from Descriptive Studies, follows with a taxonomy of types of corpora for compiling and their features and introduces the main tools at use. Moreover, the second part of the chapter is dedicated to the process of construction of the two self-compiled corpora used in this research, from data acquisition, to pre-processing procedures and compiling and designing process. The choices made and their consequent limitations will also be discussed.

Chapter 3 features the empirical analysis of the data collected through the presented corpus-based approach. All four corpora (English and Italian) are subject to wordlist and keyword

analysis in order to quantitatively calculate text reduction rates and the possible reasons behind the phenomenon in both series. Subsequently, a more detailed analysis of systematic translation strategies adopted in the rendition of profanities, idiomatic expressions and British realia in *Fleabag* is carried out through basic and parallel concordance queries. Similarly, marked language in the form of Yiddish realia in *The Marvelous Mrs. Maisel* is analysed in the last section of the chapter.

In the conclusive section of the research, an overall review of the results obtained will be conducted, as well as discussing issues and difficulties encountered throughout the analytical process in order to open up to possible replications of the process. In fact, as the research went by I found myself discovering more and more possibilities of analysis that could not be successfully embodied in this work. Therefore, my work resulted in mostly quantitative results that, while not providing a complete picture of the issue, lay the groundwork for possible future in-depth analysis.

CHAPTER 1 - AUDIOVISUAL TRANSLATION AND RENDERING SUBTITLES

1.1 Audiovisual Translation Studies: History and Development

Despite its current relevance, for decades, audiovisual texts had been considered inferior to literary text and, therefore, excluded from the discipline of Translation Studies (Perego, 2014). In fact, while the earliest AVT research dates back to 1932, Audiovisual Translation Studies were fully integrated as part of the discipline of Translation Studies only in the 1980s. However, starting from the 1990s, the consistent integration of audiovisual material into everyday life has made it impossible to ignore the cultural and academic relevance of the discipline of Audiovisual Translation in contemporary society.

Gambier (2003) sets the starting point of the consolidated discipline in 1995, coinciding with the 100-year anniversary of cinema. The increasing interest in the area of AVT enabled it to consolidate “its very own theoretical and methodological approaches, allowing it to claim the status of a scholarly area of research in its own right” (Díaz-Cintas, 2009, p. 7). Over the years, a discussion has also been raised by scholars on the most appropriate terminology to adopt when referring to this developing area of study, with each denomination reflecting a different conceptual orientation towards the discipline. At first the discussion revolved around *constrained translation* (Titford, 1982) emphasising the limitations imposed by the audiovisual format. Before the advent of the TV industry, *film translation* (Snell-Hornby, 1988) was widely employed; however, this denomination restricted the field to cinema spaces, excluding other media. Subsequently, *language transfer* (Luyken et al., 1991) was adopted, yet this formulation seemed inappropriate, as it focused on the verbal component, ignoring the semiotic transfer of the visual and acoustic channels. Contrarily, terms such as *multimedia translation* (Heiss, 1996; Ulrych, 2000) and *screen transadaptation* (Gambier, 2003) shifted the focus on the intersemiotic nature of the medium. Gambier’s notion of *transadaptation*, in particular, placed the target audience with its cultural and linguistic background at the core of the translational process. This fluctuation in its terminology is an evident sign of the ever-changing and developing nature of the discipline, which has now led to the widely accepted expression *audiovisual translation* (Díaz-Cintas and Remael, 2007; Chaume, 2012; Pérez-González, 2014). Commonly abbreviated as AVT, it is considered the main referent for this area of Translation Studies, including film, radio, television and digital video media.

Small and large screens have become integral parts of each individual’s lifestyle, from educational and professional environments, to entertainment and free time. Moreover, the

proliferation of subscriptions to streaming platforms, such as Netflix, Amazon Prime Video and Disney+ among the most renowned, has dramatically accelerated this process. This digital turn has notably sped up the consumption of audiovisual products positioning them as one of the main mechanisms of global cultural transfer. Furthermore, mainstream streaming platforms have given the audience an agency they had never had before, not only in terms of choosing what to watch, but also in terms of *how* to watch it.

Historically, each country has developed its own tradition of dubbing, subtitling or voice-over, with its own ideological background. For instance, Italy, Spain, Germany and France have established dubbing as their traditional mode of audiovisual translation; whereas Scandinavian countries, Greece and Portugal preferred subtitling and Eastern European countries adopted voice-over. These choices were far from neutral. In Italy, for instance, the preference for dubbing was strongly related to Mussolini's imposition during the fascist regime of dubbing as the only mean of translation allowed as a way to preserve Italian nationalism and identity through one of the biggest means of propaganda of the time: cinema. This historical background demonstrates how translating is never just merely a linguistic act, but also a practice shaped by ideological implications. At present, the digitalised world has slowly started to disrupt these national traditions, exposing audiences to a much wider range of translation modes and practices.

1.2 Audiovisual Translation as a Multimodal Language

The complexity of the audiovisual translation resides in emulating a “prefabricated spontaneous mode of discourse” (Chaume, 2013, p. 105) through a complex web of verbal and non-verbal codes in compliance with time and space limitations imposed by the translation format. An audiovisual text is “a communication act involving sounds and images” (Zabalbeascoa, 2008, p.22). Consequently, audiovisual products inherently present a multimodal hybrid nature providing information through four channels of discourse operating simultaneously: the verbal acoustic channel (spoken dialogue), the nonverbal acoustic channel (sound effects and soundtrack), the verbal visual channel (subtitles, captions, on-screen text) and nonverbal visual channel (images and iconography). As a consequence, the discipline of Audiovisual Translation Studies requires an interdisciplinary approach that intersects Translation Studies and Film Studies.

Chaume elaborates further on this point arguing that film language per se encompasses a series of codes that heavily affect the translational approach (2004b, p. 16). First and foremost, the linguistic code, which is evidently shared among all genres of translations and is not

exclusive to the field of AVT, in the case of audiovisual products it presents challenges related to the rendering of a spontaneous form of speech, “written to be spoken as if not written” (Chaume, 2004b, p. 17). Beyond the verbal code, paralinguistic codes (incorporating intonation, pitch and volume) and mobility codes related to proxemics signs and characters’ mouth movements are fundamental when dealing with the synchronisation of dialogues and in compliance with the delivery of the actor. Similarly, musical and sound effects codes contribute to the narrative atmosphere, hinting at how emotions must be decoded. Furthermore, iconographic, photographic and graphic codes should not be underestimated in translation, especially when dealing with marked language and idiomatic uses of language that carry a series of imagery and meanings that need to be translated remaining coherent with the iconographic aspect on screen. Lastly, all of these distinct codes are related to one another and connected through the editing process with its own syntactic code and cinematic punctuation marks (fade-outs, juxtapositions, cuts, dissolves).

These multimodal implications reinforce Gambier’s focus on the intersemiotic nature of the discipline, whose ultimate goal remains accessibility (2003, p. 179), a term that, according to the scholar, encompasses a series of parameters. On a linguistic level, it includes the acceptability of the language with its own style and norms, as well as domestication strategies to smoothly incorporate the “otherness” of cultural elements. From a technical perspective, accessibility also involves legibility and readability parameters of the text on screen, synchronicity of the subtitles to the visual elements and the management of relevance of which information to retain or omit. Among these codes, the linguistic component remains of particular relevance for the present study. Unlike traditional written genres, the language of audiovisual products simulates spoken discourse, it is paradoxically “scripted” to sound spontaneous, creating a variety of language that Chaume and Baños-Piñero define as *prefabricated orality*: “creating fictional dialogues that sound natural and believable” (2009, p.1). This phenomenon implies a diamesic shift, through which distinctive features of spoken language are literally scripted to construct the story and the character’s identity.

Overall, audiovisual translation research analyses audiovisual products not only in terms of linguistic grammar, but also of cinematic grammar, characterised by its own features and conventions. As the present research is mainly concerned with the linguistic and translational dimension of the discipline, this cinematic aspect will not be examined in detail; though, it should be briefly acknowledged due to its influence on the subtitled text.

1.3 Modes of Audiovisual Translation: Revoicing and Captioning

Audiovisual translation modes include all types of approaches to transfer audiovisual texts, whether intralingual, that is to say within the same language, or interlingual, between two languages and cultures. This discipline relies on a macro-distinction of two categories of modes of AVT: *revoicing* and *captioning* (Chaume, 2013). The difference between the two approaches relies primarily on the prioritised semiotic channel: the former operates on the acoustic track, manipulating or replacing it; the latter works on the visual channel, introducing written text onto the screen. As previously noted, historically, each country has privileged one mode over the other, distinguishing dubbing countries (Italy, Spain, France and Germany) from subtitling countries (Portugal, Greece, Netherlands, Scandinavian countries).

However, with the unstoppable development of new technologies and streaming services, alternative and more accessible modes of subtitling and dubbing have been introduced in order to meet the needs of different social groups (Chaume, 2013, p. 107). Today, captioning includes specialised modes of translation such as Subtitling for the Deaf and Hard of Hearing (SDH), live subtitling, surtitling for opera and theatre and fansubs, as well as fandubs and Audio Description (AD) for the visually impaired in the case of revoicing. Each of these practices comes with its own set of peculiarities and limitations. The present study, while theoretically treating both modes of AVT, focuses the empirical analysis exclusively on the captioning modes.

1.3.1 Revoicing: Dubbing and Voice-over

The revoicing mode of Audiovisual Translation consists in producing and recording a new soundtrack and subsequently synchronising it to the original video, either deleting or lowering the original track. The two standard forms of revoicing are dubbing and voice-over. Dubbing consists in “translating and lip-syncing the script of an audiovisual text” (Chaume, 2013, p. 107) and it is the most commonly used mode of audiovisual translation in European countries (such as Italy, Germany, Spain and France), Turkey, China and Japan. Voice-over or partial-dubbing, on the other hand, consists in playing the newly recorded track in the target language, as the original dialogue plays at the same time in the background with a lower volume so that the translated message can be easily heard by the audience. In order to do so, the target voice usually enters a few seconds after the beginning of the original sentence, so that the audience can have an idea of what the original voice sounds like. It is most commonly used for documentaries and news broadcast all over the world, and for films and TV shows in countries such as Poland, Russia, Estonia and Latvia. In these areas, the choice of keeping the original

voice in the background is perceived as a more credible mode of translation, more faithful to the source.

Voice-over is at times called narration, when the voice playing over the original track does not utter a literal translation, but rather a condensation of the original dialogues, as is common in educational historical documentaries. Another mode of revoicing that drifts even further from faithful renditions of the source material is free commentary. In this form of adaptation commentators not only can manipulate and condense the text, but they can even add information absent in the source text and personal opinions, as it is typically the case in sports commentary. Other modes of revoicing include simultaneous interpretation of film, now in disuse, in which an interpreter was present as the film was shown on screen to translate (or better, to interpret) the original utterances in real time. Historically, this represented one of the first forms of dubbing employed to cater for the advent of sound in film. This practice is still at use in Thailand and, on rare occasions, within international film festivals. Finally, the request for more accessibility in the field has led to the introduction of the practice of Audio Description (AD) for the blind and the visually impaired. Audio Description is not limited to rendering the verbal dialogues, but it also provides further detail of what is happening on the screen, transforming visual and graphic signs into words. Recently, this practice has been also implemented in museums and cultural heritage sites, demonstrating the universality and inclusivity of this instrument.

1.3.2 Captioning: Subtitling

Within an increasingly interconnected global context, language transfer becomes pivotal. The most widespread mode of AVT is captioning, more specifically subtitling. In their volume, Díaz-Cintas and Remael (2007, p.8) provide a definition that captures the semiotic nature of this practice, described as:

“a translation practice that consists of presenting a written text, generally¹ 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 and the information that is contained in the soundtrack”.

¹ In some cases, such as the Japanese language, subtitles can also be presented vertically on the right side of the screen.

As underlined by this definition, subtitling is a multifaceted operation that does not merely transfer words, but creates a new verbal layer, which has to be coherent with both the acoustic and visual channels in the source text. Generally, subtitling practices are considered cheaper than dubbing and have the benefit of allowing the audience to listen to the original voices of the interlocutors and enjoy the original performances in their entirety.

Gottlieb defines subtitling as an “overt type of translation” (1994, p. 102) that, retaining the original acoustic track, remains open to criticism from anybody who even partially understands the original utterance. In contrast, he also refers to subtitles as “fragmentary”, indicating their ability to render only the lexical and syntactical features of the dialogue, leaving aside the prosodic features of dialogue. At the same time, subtitles can also represent a source of distraction for the audience: when confronted with both textual and visual input, the human eye tends to prioritise the verbal component². Despite these challenges, in recent years, subtitling has slowly become more and more well-known even in European settings where dubbing had historically been perceived as the preferred mode of AVT. This is also a pedagogical choice made when dealing with cartoons or programs for children, as a way to naturally expose younger audiences to foreign languages and improve their learning experience.

As mentioned before, the history of subtitles dates back to early silent films with intertitles or title cards, that is to say brief textual inserts displayed on screen between scenes, usually on a black background, to convey dialogue or relevant narrative information. These used to be easily translated and replaced with the target language version, giving rise to one of the first instances of audiovisual translation in cinema. In other cases, a live narrator or interpreter was present in the theatre where the film was screened to translate these insertions to the audience in real time. Nowadays, theatres and opera productions use a specific variant of subtitling, called surtitles, which are displayed on a dedicated screen positioned above the stage that help the audience understand the dialogues and the storyline without any text occupying the visual performance on stage. Another result of technological advancement has been the introduction of re-speaking or live subtitling. This technique is used for live TV broadcasting that requires a voice recognition software that simultaneously reproduces on screen a summary of what a professional interpreter produces and speaks into the microphone connected to the speech recognition software.

² Gottlieb argues that for this specific reason, when dealing with informative programs such as documentaries, revoicing strategies seem more adequate (1994, p.103).

For pre-recorded programs, Subtitling for the Deaf and the Hard of Hearing (SDH) or closed captions (CC) is another mode of translation that enhances accessibility and compensates for the total or partial absence of sound. The main difference from standard interlingual subtitles lies in specific formatting and technical conventions. In fact, colour coding used to be adopted to differentiate between speakers and signal the turn-taking system in the conversation; now, this practice is slowly becoming obsolete. Paralinguistic elements are described in square brackets and capitalised (e.g. [*WHISPERING*]) as well as background sounds (e.g. [*DOOR SLAMS*]), songs (e.g. [*POP MUSIC*]) and accents (e.g. [in an Italian accent]). Furthermore, these subtitles stay on screen for a longer time and can be up to four lines long compared to the standard conventions of the format (Zabalbeascoa, 2025, pp. 80-82). Finally, another mode of audiovisual translation that gives agency to the audience is fansubbing, that is to say homemade fan-produced subtitles for television series, films and animes. Given their divergence from mainstream industry constraints, specialised corpora of fan-produced subtitles constitute the core of the present investigation exploring this practice and its implications.

1.3.3 AVT: Professional vs. Amateur Translators

In this increasingly democratic digital era, the traditional lines between professional and amateur translators have been blurred by the rise of the well-known practices of fansubbing and fandubbing. Technological advancements and computer subtitling programs³ becoming user-friendly have rendered this practice easily accessible and affordable by anyone who is passionate about the field.

Fansubbing as a practice dates back to the 1980s as a response to the spreading of Japanese culture, especially *anime*, and turned into a mass phenomenon with an immense virtual community behind it. Passionate viewers started translating these audiovisual products themselves in order to overcome the linguistic barriers, in all likelihood violating several copyright regulations. As a result, these fan-generated subtitles were distributed freely over the Internet for fellow enthusiasts. Despite the non-commercial voluntary nature of the practice and the amateurish consideration, Díaz-Cintas demonstrates (2006, pp. 38-39) that even fansubbing groups operate as a proper team of “professionals”. Within these groups, each member covers a specific role: raw providers or rippers provide the source video material to translate; proper translators produce the linguistic transfer; timers deal with the cueing procedures; typesetters

³ Aegisub and Subtitle Edit are the two main subtitling softwares adopted by amateurs, they are open-source, user-friendly and free for download.

choose the formatting (fonts, colours and screen positioning); editors and proof-readers refine the coherence of the final work and encoders compress the final digital product.

However, while professional subtitles are rigidly bound to the standards of the format, these fansubs have always appeared to be more creative and less inclined to these limitations. Fansubbers generally tend to stay closer to the original text and its specificity as they are fanatics of the original product to begin with and could, therefore, be considered more inclined to source-oriented (Venuti, 1995) renditions. Consequently, they often overlook subtitle conventions by using glosses and notes, usually at the top of the screen, to explain opaque references and, consequently, challenging the traditional tendency of subtitles to be as invisible as possible. Furthermore, as fansubbers are not necessarily certified professionals in both source and target languages, translation errors could be found in their renditions. These inaccuracies could be due to the lack of fluency in the language, as well as to the rate at which they are produced to ensure their immediate availability, despite the lack of financial compensation for their voluntary contribution to the virtual community they are part of. Overall, fansubbers tend to merge conventions from TV, cinema, subtitling for the hearing impaired and localisation, giving rise to hybrid forms of subtitling applying several translation strategies and constantly evolving as the digital revolution continues.

“Fansubs share some of the characteristics of professional subtitling but they are clearly more daring in their formal presentation, taking advantage of the potential offered by digital technology” (Díaz-Cintas & Muñoz Sánchez, 2006, p. 51).

Analogously, fandubbing (or *fundubbing*) has also found its place in dubbing TV shows, cartoons and film trailers with a microphone at home. However, this procedure often falls into parodic versions in which the original script is manipulated.

In the present-day context, the role of fansubbing has evolved significantly with the popularisation of streaming platforms. As mentioned before, in the early days, the rise of this practice was related to filling a gap in the market: translating content that was not accessible in any other way. At this point in time, platforms publish official subtitles almost immediately in multiple languages, reducing the demand of fan-produced alternatives. Nevertheless, fansubbers continue to work on unavailable content and on producing alternative translations for material whose official subtitles are considered inaccurate or not faithful enough to the source text, especially in complex genres like contemporary comedy.

1.4 Subtitling: Technical Constraints and Issues of Fidelity to the ST

Interlingual subtitling as an audiovisual mode is fundamentally driven by the principle of synchrony: image, sound and written text must appear on screen at the exact same time and long enough for the audience to be able to read the dialogue and understand its meaning, which has to semantically reproduce the original (Díaz-Cintas and Remael, 2007, p. 9). The co-existence of these two codes (image and sound) and their rigid synchronisation involves a series of technical issues, time and space parameters and constraints, which pose challenges for translators and influence their linguistic choices.

First and foremost, space constraints are defined by the screen frame. Historically, they imposed a number of 32 to 37 (maximum 41 in rare occasions) characters per line (CPL) in a maximum of two lines for TV programs for the Roman alphabet; in proportion, one line normally rendered a three-second utterance. Conventionally, subtitles are positioned horizontally at the bottom of the screen aligned in the centre, to minimise visual disruptions⁴: they can be pyramid-shaped (the bottom or the top line is longer than the other), or square-shaped (both lines are the same length) (Zabalbeascoa, 2025, p. 78). In terms of typography, subtitles are usually white or, more rarely, yellow, with a thin black shadow or contour that ensures legibility on any kind of background.

As for temporal constraints, cueing or spotting determines when a subtitle should appear and disappear from the screen and how the dialogue should be divided into units, in compliance with the ideal synchrony with the original utterance. In order to calculate the time and speed a subtitle can stay on screen, the six-second-rule is employed. This formula refers to the time it takes the average viewer to read, comprehend and assimilate the message on screen. According to calculations, the average viewer can read a maximum of 70-74 characters in six seconds. If periods are longer, they should, therefore, be split into smaller units of text (Ivarsson & Carroll, 1998, p.74). In addition, from 2 to 4 frames should be left as a pause between consecutive subtitles in order for the reader to register the change (Díaz-Cintas, 2006, p. 95). To be more accurate, this rule is mostly used in television due to its heterogeneous public, whereas for other media such as cinema and Internet, a generally faster reading speed is usually taken into account. For instance, cinemas and DVDs can use up to 40-41 characters per line: the former due to the higher definition on screen, the latter as the receiver can easily go back and replay a

⁴ On some occasions, subtitles can be moved to the top area of the screen if the background of the bottom area makes them not readable or if there is something else written at the bottom (e.g. the names of the actors).

specific part to rewatch. Furthermore, cinema tends to use more line breaks so as to keep the text at the centre of the screen.

Beyond these limitations, one of the main features of the target text should be semiotic cohesion (Chaume, 2004b, p. 23); when semiotic cohesion is guaranteed, a gap in the subtitled version of the dialogue can be immediately filled with the visual on screen providing the viewer with the information missing from the text. The visual, together with the narrative context, supports the coherence of the rendition in the target language, even with some level of loss in the translation. In order for this to be assured, once again synchronicity comes into play: if the subtitles are perfectly in time with the oral utterance and the visual appearance, cohesion is preserved. Modern advanced subtitle programs are provided with a speech recognition tool that discerns the exact timing of the utterance. Moreover, images can, in turn, be a useful tool for the subtitlers to rely on in order to avoid redundant information that can be easily acquired from the visuals and could be omitted or condensed in compliance with the spatial constraints.

In summary, when rendering their translated dialogue, subtitlers depend on the intersemiotic channel to convey the message, as the image can show what words cannot. Despite fidelity being one of the aims to strive at when subtitling, the translator can remain faithful to the original only to a certain extent, not only due to these technical restrictions, but also to the unpreventable cultural and linguistic differences from the source.

1.4.1 Subtitling: Punctuation Conventions in AVT

Besides spatial and temporal constraints, general guidelines can also be identified for punctuation. Discussing proper universal norms for punctuation in subtitling would be complex, as its style is highly dependent on the type of product; however, a series of standard conventions are deemed more acceptable than others. Overall, it could be said that punctuation in audiovisual products preserves standard grammatical functions and conventions. In some instances, however, its function can vary according to the format; here, a brief list of some of the most common instances will be reported.

The use of commas, for instance, does not necessarily respect standard grammatical rules, but rather becomes a useful device to render complex sentences simpler and guide the audience's reading velocity. Furthermore, when positioned at the end of a subtitle line, it signifies that the utterance is incomplete and will continue in the following line of subtitle. As concerns dashes (or hyphens), in subtitling grammar, they are used as a marker of turn-taking in dialogues, especially when dealing with fast-paced interactions. When placed at the beginning of each speaker's utterance, dashes indicate a change in the speaker when two lines

of dialogue appear together on the same frame. Suspension points (or three dots) can be used at the end of a sentence, as with commas, to indicate that the utterance continues in the next line and, correspondingly, at the beginning of the sentence to signal the continuation of the previous line. Furthermore, they can also be used to codify pauses, hesitations or interruptions in the dialogue. This specific expressive function serves as a point of departure for the upcoming analysis in this research. Suspension points and dashes will be closely examined as a graphic device for rendering speech fragmentation in *Fleabag*.

Beyond punctuation marks, other typographical conventions in subtitling involve italics, a mechanism used to highlight a specific term in the utterance (sometimes substituted by quotation marks), as well as to account for voices in the background or off-screen dialogue. It is also used for utterances coming from a technological device, such as TV or computer, or even a person on the other end of a phone call. Moreover, even though it is not recognised as an official convention, italics is more and more often used as a marker for neologisms, words that do not belong to the target language vocabulary, literary references, film and song titles, as well as when a foreign language is spoken in the scene.

1.5 From Spoken to Written Dialogue: Diamesic Shift and Prefabricated Orality

Subtitling always implies a diamesic shift, that is to say a change in medium, from oral to written speech, which heavily influences the translation process. Henrik Gottlieb (1992, 1994) conceptualises interlingual subtitling not as a linear linguistic transfer, but as an intrasemiotic and diagonal translation: intrasemiotic, as it operates within the limits of film and TV media; diagonal because the translator must simultaneously perform a transition between two different cultural languages (horizontal transition) and from spoken foreign sub-code to written domestic sub-code (vertical transition). Furthermore, Gottlieb breaks down the fundamental differences between natural speech and standard written prose. Firstly, in spoken dialogue the interlocutors naturally share a situational context in which details about the communication are implied, whereas in written texts, the target can be unknown and the message requires to be fully explicated. Moreover, formal and informal parameters in spoken language are less rigid than those of written discourse. In addition, the (artificial) spontaneous speech found in audiovisual media presents a series of pauses, false starts, interruptions and self-corrections, as well as nonsense, ambiguous situations and overlapping of voices; in some instances, it can also feature dialectal expressions or regional accents.

As mentioned before, audiovisual texts, and more specifically subtitles, by nature, feature a type of discourse that is designed to appear spontaneous and natural, as a faithful

reproduction of spoken discourse, while actually being *prefabricated*, using the term coined by Chaume (*oralidad prefabricada*, 2009). As a matter of fact, screenplay dialogues are far from spontaneous; rather, they are scripted to be performed as if they were naturally spoken and are actually governed by a series of linguistic norms aimed at mirroring spoken discourse, yet fundamentally belonging to the written domain. For instance, as posited by Díaz-Cintas and Remael, standard conversational features such as hesitations, repetitions and false starts are often omitted from subtitle lines unless they hold a narrative or character-defining function (2007, p. 63).

While subtitling varies from genre to genre, some universal features can be easily identified. Overall, the main strategies of translation show signs of text reduction. Díaz-Cintas and Remael (2007, p.146) justify this systemic reduction stating that the human brain can process speech faster than it can read and absorb written text; in addition, the audience must simultaneously take time to watch the scene on screen, listen to the soundtrack music and evidently combine the different modes of this language. The two main strategies deployed to put text reduction into practice are partial reduction and total reduction. The former is carried out through condensation strategies involving a syntactic reduction or a simplification of the source utterance so as to avoid redundancy, as well as through reformulation, which involves a rewriting of the original utterance that preserves the core meaning while optimizing reading speed. Total reduction, in contrast, manifests itself through omission or deletion strategies involving a radical elimination of lexical items, phrases or sentences from the target text in compliance with the temporal and spatial restrictions.

One of the operational principles that guides the subtitler's choice of translation strategy is the Principle of Relevance. Salvatore Attardo argues that the process of translation is approximate by nature, indeed, as a perfect translation does not exist, neither intralingual nor interlingual⁵ (p. 174). On the basis of this assumption, the best solution needs to be considered: the question the subtitler must ask is which items are essential to preserve in the target text and which can be omitted, according to visuals, acoustics and the general context of the dialogue. In this transition from acoustic to written form, some of the features of simulated spoken language end up inevitably lost in translation. Consequently, the issue of *naturalness* comes into play, as mentioned by Pavesi (2019) in her corpus study of audiovisual translation. The concept of "naturalness" was firstly introduced by Romero-Fresco when referring to a "nativelike selection of expression in a given context" (2009, p. 51). Pavesi (2019, p. 324)

⁵ Jakobson's terminology (1959) distinguishes between intralingual, interlingual and intersemiotic translations.

further provides a taxonomy of linguistic features to evaluate naturalness in dubbing, which prove to be equally relevant and applicable to the present study of subtitling. Among these features, she refers to discourse markers, greetings, intensifiers, hesitations, vocatives and slang language. Indeed, a category that is most frequently omitted in subtitling is that of discourse markers. Despite lacking a denotative or narrative function, these particles are indispensable for reproducing the architecture of spontaneous dialogues and to convey the logic behind coherent conversations and spoken discourse (Chaume, 2004a, p. 844). As the translator's main goal should always be to produce the same perlocutionary effect on the target audience as the source text, the systematic deletion of these particles heavily compromises the naturalness of dialogue.

Schiffirin defines discourse markers as cohesive devices reflecting connections between propositions. Moreover, they present several properties: they are multifunctional, they are not obligatory, even though their absence can be noticed, and they are diverse from a syntactic point of view (1987, p. 64); furthermore, they help the interlocutor's interpretation of the message in a conversational context. She categorises discourse markers (*oh, well, and, but, or, so, because, now, then, you know, I mean*) according to their function. Based on her framework, *oh* is considered a marker of information management with both productive and receptive tasks. It is usually considered as an exclamation or interjection signalling a sudden change in the speaker's cognitive state, due to surprise, fear or pain. It can also be used to preface questions, answers and acknowledgements, to integrate a piece of information, to recognise familiar information or when the information provided does not correspond to the speaker's assumptions. *Well* is categorised as a response marker that accomplishes coherence in a conversational exchange, but has no referential meaning, whereas *and, but* and *or* are discourse connectives with a coordinative role. *So* and *because* are discourse markers conveying cause and result, *now* and *then* are temporal adverbs marking a speaker's progression in time. Finally, *you know* (or *y'know*) marks transitions in information and *I mean* marks the speaker's orientation towards modifying ideas and intensions.

When moving from the spoken dialogue to the constraints of the screen, however, these discourse markers represent the most vulnerable category during the process of text reduction (Díaz-Cintas & Remael, 2007). This tension will occupy a central position in the analytical chapter of the present research. In contemporary comedies characterised by rapid-fire conversations and fragmented speech, the omission or preservation of discourse markers and slang cannot be justified as a merely technical choice; on the contrary, it heavily influences the original comedic effect, the authenticity of the dialogues and the characters' identities.

1.5.1 Marked Language in Subtitling

In addition to managing text reduction due to the technical parameters, subtitlers might simultaneously deal with non-standard linguistic varieties of speech within the audiovisual text. In this context, marked speech is:

characterized by non-standard language features or features that are not “neutral”, even though they do belong to the standard language, and may therefore have more or less specific connotations. Speech can be marked by style or register, and it can be either idiosyncratic or bound to socially or geographically defined population groups. Besides, marked speech includes taboo words, swearwords, and emotionally charged utterances such as interjections and exclamations.

(Díaz-Cintas and Remael, 2007, p. 187)

Nevertheless, a difficult challenge for the subtitler is that of preserving the character’s peculiar manner of speaking across several linguistic dimensions, particularly style, register and the use of dialect and slang. Style refers to an individual’s manner of expressing oneself, whether speaking or writing. Given the assumption that ideally subtitlers should attempt to retain these stylistic features, when they cannot be directly reproduced in the target text, it is common for the subtitler to resort to compensation strategies somewhere else in the utterance. Register, on the other hand, refers to the variety of language and the degree of formality employed in a specific social context (Díaz-Cintas and Remael, 2007, p. 189). Register can often be an indicator of a character’s social status, as well as hierarchies in certain relations.

Dialect and slang are characterised by non-standard grammatical features and marked lexical choices, as well as peculiar phonological traits. Dialects, in particular, refer to variations in pronunciation in comparison to the Standard English spoken by the majority of the English population, associated with a specific geographical area. Nevertheless, this doesn’t imply that dialects are incorrect ways of speaking language (Ranzato, 2010, p. 109). However, they are bound to the geographical area they are spoken in; therefore, it is unlikely for the target language to present a direct equivalent. Even when translators opt for translating one dialect with another one considered to have a similar connotation in the target language and culture, this process remains problematic, as dialects are not simply linguistic variations, but imply a strong cultural identity. On the other hand, slang functions as a non-standard variety often resembling “anti-language” in which conventional lexical choices are substituted with informal alternatives tailored by a social group (Díaz-Cintas and Remael, 2007, p. 191).

In the practice of subtitling, grammatical mistakes and dialectal markers tend to be corrected and neutralised. Accents and pronunciation variations are difficult to render onto the written line, despite being crucial for the linguistic and social identity of the characters. To address these constraints, a corpus-based approach proves useful for analysing the linguistic “markedness” of the two selected audiovisual products evaluating how idiomatic and slang expression are translated into the target text through diverse translation strategies.

1.5.2 Slang and Profanities

A particular dimension of marked dialogue involves the use of slang and profanities. Within this context, Díaz-Cintas and Remael (2007, p. 196) remark that taboo words and swearwords are often toned down in subtitles or even omitted, as a direct consequence of technical constraints. However, these expressions are emotionally charged and perform a pivotal phatic function, as well as reflecting the cultural identity of a community. As a consequence, subtitlers must carefully assess their communicative value within the source text in order to select an appropriate equivalent in the target language.

A taboo word may be defined as a lexical item whose use is restricted or prohibited by social customs, whereas swearwords are offensive words conveying anger, frustration, or excitement. These two categories can often overlap, as swearwords can on some occasion become taboo language, and they both posit a challenge in their rendition. Therefore, attitudes towards offensive language and profanities vary significantly across cultures; as a consequence, translators need to evaluate their acceptability in the target text when considering how to render them. Even widely recognised and used expressions can become problematic outside their original contexts of use, as the meaning of profanities is highly context-dependent. The occurrence of the swearword *fuck*, for instance, may convey anger, emphasis, surprise, disgust or even amusement based on its context of use and on the speaker’s intention. As a result, rather than a simple literal transfer, what translators should really take into consideration is the emotional effect these expressions produce and whether they influence the narration.

Finally, another matter that deserves careful consideration when translating marked language in subtitling is the intersemiotic shift from oral to written language. In fact, written discourse traditionally has lower tolerance for profanities compared to spoken dialogue. This difference implies further questioning on the part of translators on how to render these elements preserving their acceptability in the target context.

1.6 Translating Culture: Realia and Culture-Specific Items

To navigate the web of constraints caused by the diamesic shift, subtitlers must balance technical boundaries and cultural fidelity. As Gottlieb (1992, p. 162) notes, subtitling has been historically regarded as a “*necessary evil*” in screen translation. Not only this mode of translation is written, it is also additive, as verbal material is added onto the original work; immediate, because its speed is beyond the audience’s control; synchronous, as target written text and source dialogue track are presented at the same time; and polymedial, as it employs at least two semiotic channels to convey its message. While his pioneering classification remains an interesting reference as it focuses primarily on measuring how much text is omitted or condensed, the present study will adopt the framework introduced by Díaz-Cintas and Remael (2007) as the primary analytical tool. Their taxonomy specifically addresses the management of cultural specificity, highlighting procedures based on how the cultural weight of the original text is negotiated for the target audience.

Indeed, the challenge of translators is not restricted to mechanically transferring the meaning of words from one language to another; most importantly, translators work as cultural mediators between the source and target culture (Pedersen, 2010, p. 67). Understanding when to put this mediation into practice constitutes a serious matter for these professionals; for subtitlers, it is an even more relevant issue as audiovisual material is typically pervaded by culturally marked items, often visually present on screen and impossible to ignore. When a cultural item is unknown to the audience, there is little a translator can do with the limitations of the medium. According to Arias-Badia (2015, p. 513), there is a general tendency on behalf of translators to resort to more neutral occurrences in order to make the text more comprehensible to the target reader. The main challenge remains finding a balance between remaining faithful to the source culture and its peculiarities and adhering to time and space restrictions. For instance, when a cultural item appears in the text, translators should ask themselves how relevant that reference is in the source text and whether it is equally recognisable in the target culture to such an extent that it can be retained as it is in the target text, without further intervention. Pedersen (2010, p. 76) suggests the use of corpus research as a useful tool to investigate whether a term is culturally specific and how it is perceived within the target culture.

According to Díaz-Cintas and Remael, “culture-bound terms are extralinguistic references to items that are tied up with a country’s culture, history, or geography, and tend therefore to pose serious translation challenges” (2007, p. 200). These culture-bound references are also called *realia*, in Florin’s words:

Realia (from the Latin *realis*) are words and combinations of words denoting objects and concepts characteristic of the way of life, the culture, the social and historical development of one nation and alien to another. Since they express local and/or historical color they have no exact equivalents in other languages. (1993, p. 123).

It becomes clear that rendering realia does not imply a simple transfer of words, but rather of concepts and cultures and a universal solution cannot be expected. However, a categorisation of these culture-bound terms can help translators identify their function in the text and find the most suitable solution to preserve their otherness. Florin (1993, p. 123-124) distinguishes three main categories of realia, which will be used in the present study to examine the selected audiovisual products: socio-political realia, related to social groups, sub-groups and institutions, geographical realia (which can be regional or international) and ethnographic realia. This final category is the most heterogeneous as it encompasses everyday life, work, art and religion. In the present study, which dedicates a section to the translation of Yiddish cultural items, ethnographic realia play a pivotal role. Specific terms carrying strong religious and cultural meaning, such as *mitzvah*, *mezuzah* or *hannukah* will be closely examined, as they cannot be rendered without some degree of manipulation.

When dealing with these realia, the question of acceptability becomes crucial, as some of them can be easily accepted in the target language. When they are not considered acceptable in the target text, the translator needs to resort to translation strategies. Different scholars have tried to propose a list of the dominant strategies adopted by translators to render the cultural specificity of realia. Díaz Cintas (2003) and Santamaria Guinot (2001) differentiate between the following strategies reported in Díaz-Cintas and Remael (2007, p. 202): loans, in which the source text is transferred as it is in the target language, preserving its cultural specificity through a foreignising approach; calque (or literal translation), in which the expression is translated word for word⁶, even if there might be a risk of sounding unnatural. Explication strategies, on the contrary, avoid opacity in the target text through generalisation, hyponyms or hypernyms, that contribute to the loss of specificity, yet make the reference more transparent to the audience. A variant of this strategy commonly used in subtitling, is substitution, through which a source-language element is replaced with a functionally equivalent target-language element;

⁶ On some occasions, these terms might still require an explanation, which can usually be given in brackets. Clearly, in the case of subtitling, this option is not contemplated due to its constrained space.

in these instances, space constraints are respected and the perlocutionary effect is preserved. When no room is given for explicitation and the target audience is not familiar with the culture-bound term, translators resort to transposition strategies. On these occasions, a cultural concept from the source language is replaced by one from the target language; however, this strategy can become problematic if the visuals provide a clear reference for the original realia for which there is no correspondence in its target rendition.

There can be occurrences in which the translator's creativity must be put to the test and lexical recreation is necessary to reproduce the neologism or wordplay in the original text. This practice is particularly useful when dealing with humoristic texts. However, some culturally-specific terms remain impossible to translate, even partially. That is when compensation strategies must be employed in order to retain the perlocutionary effect in a different part of the text, paying attention to its coherence with the visual code. When even this kind of strategy cannot be put to use, omission is the last resort: the culture-bound item is removed completely, either due to space limitations or to the absence of a corresponding reference. Finally, although rarely, additions of information can also occur in subtitling when the cultural reference is extremely relevant for the narration.

When a culturally marked text is transferred from a source to a target environment, the inevitable cultural asymmetry confronts the translator with a serious challenge. Javier Aixelá provides his own definition of culture-specific items:

Those textually actualized items whose function and connotations in a source text involve a translation problem in their transference to a target text, whenever this problem is a product of the nonexistence of the referred item or of its different intertextual status in the cultural system of the readers of the target text.

(1996, p. 58)

Moreover, Aixelá organises translation strategies into a taxonomy based on the degree of intercultural manipulation involved and distinguishes two macro-categories. The first category encompasses translation strategies of conservative nature, which prioritise a foreignising approach and preserve the "otherness" of the source text. Within this group, repetition occurs when the original cultural item is transferred as it is into the target text; here, while preserving the cultural authenticity, it risks sounding opaque to the target culture. When the original reference is expressed using a different alphabet, translators resort to orthographic adaptation through procedures of transcription or transliteration. A more accessible approach is conveyed

through linguistic (non-cultural) translation, where the translator chooses a target-language close reference to the original item that appears more comprehensible while still being part of the ST cultural system. Finally, Aixelá identifies two forms of glossing: extratextual, which explains the rendition using tools such as footnotes, glossary, commentaries or brackets, and intratextual, through which the explanation is incorporated in the text so as not to disturb the receiver's attention. Within subtitling, however, these two strategies appear inadequate due to the constrained nature of the audiovisual medium.

The second category comprises translation strategies of substitutive nature, which lean on more domesticating approaches and are more target-oriented. Through synonymy, the translator opts for a synonym to avoid repetitiveness. When a culture-bound item is deemed too obscure for the target audience, the translator resorts to limited universalisation, replacing the source reference with a less specific, but more accessible reference belonging to the same culture; if translators cannot find a suitable solution, they may opt for absolute universalisation, translating the realia into a generic and neutral equivalent. Through naturalisation, the culture-specific item can be completely adapted into the target culture, even though, due to the visual aspect, it rarely fits the rendition of audiovisual products. When the cultural reference is considered unacceptable or non-essential to the plot, subtitlers can adopt deletion strategies, omitting it entirely. On the contrary, on rare occasions, translators can let their creativity out through autonomous creation, introducing a cultural reference that does not exist in the original text. Finally, Aixelá mentions three other valuable options: compensation, which compensates a cultural loss by introducing another cultural marker somewhere else in the text to reproduce the perlocutionary effect; dislocation, involving a displacement of the cultural reference in a different part of the target text, and attenuation, which involves replacing it with a more socially acceptable reference.

The selection of one strategy over the other is heavily influenced by a series of variables discussed by Florin (1993, pp. 126-128). First, the choice depends on the general features of the text, such as genre and communicative purpose. Second, the translator must evaluate the importance of the realia in context and its intrinsic nature, determining its relevance and acceptability into the target text. Third, peculiar features of the source and target texts might influence these choices, and, finally, the expectations, cultural background and knowledge of the receiver. In the context of this research, Florin's criteria provide a framework to contextualise the translation choices analysed, as they are not arbitrary decisions, but systematic solutions to the pressure of preservation of text and contextual features of the original elements.

1.7 Rendering Humour: the Issue of Untranslatability

When dealing with the audiovisual translation of contemporary comedies, not only do translators have to consider the limitations of the audiovisual medium, but they also have to confront the challenges related to the humoristic features of language. The debate surrounding the translatability or untranslatability of humorous language continues to occupy a central position in several scholars' works. Rendering humour has often been deemed problematic due to a combination of rigid linguistic boundaries, social norms and cultural specificities. As a matter of fact, it is undoubtable that culture-specific allusions can be particularly difficult to render in a different culture and the linguistic ambiguities and wordplays of the source text can seem impossible to preserve in the target language. Yet, not only is humorous translation possible, but by identifying the linguistic mechanism used in the source text, it can at times generate an even stronger effect on the target receiver. Nonetheless, a certain degree of loss and manipulation is inevitable: it can be technical, when justified by technical constraints, or operate on ideological or functional levels (Dore, 2020 p. 40).

According to Attardo, finding the right definition for humour in a general sense is impossible, as it has been looked at from different points of view and theoretical approaches. He affirms that scholars treat humour as a broad category encompassing all events and objects perceived as funny and eliciting laughter⁷ and amusement (1994, p.4). Furthermore, he distinguishes three main branches of humour theories: superiority theories drawing on Plato, Aristotle and Hobbes's implication that people find amusement in laughing at other people when they feel superior to them; incongruity theories, based on Schopenhauer and Kant's consideration of humour as an expression of opposition between meanings and expectations; and relief theories, reprising Spencer and Freud's vision of laughter as a manifestation of pleasure. It is evident that humour involves more than simple words. It is, indeed, a complex matter of culture and ideology: not every culture, in fact, perceives humour in the same way, and in several instances, translating humour can become ethically problematic, other than extremely challenging. When faced with the issue of translating humour in audiovisual products, the standard challenge lies in the fact that humour normally resorts to wordplay, puns, ironic and witty comments. Here, the translation process becomes more demanding as it is often immediately linked to the visuals on screen, which cannot be manipulated in any way.

⁷ Attardo remarks that laughter does not directly equal humour for different reasons: laughter can be provoked by other emotions, such as happiness or embarrassment, and its intensity does not signal the success of the humoristic effect, which might only solicit a smile and still be functional.

In order to navigate this intricate road, Zabalbeascoa (1996, pp. 251-255) creates a “map of humour”, that is to say a classification of different types of humour and jokes that translator can benefit from when considering which strategies to adopt. The main categories identified by the scholar are: international or bi-national jokes, in which the humorous effect does not rely on language- or culture-specific references and, in most cases, can be translated literally; on the contrary, national-culture-and-institutions jokes are strongly culture-bound and require a form of adaptation to preserve the humorous effect for the target audience. Similarly, national-sense-of-humour jokes reflecting a community’s sense of humour and are often inside jokes made at the expense of a sub-community and the adaptation can be controversial. Moreover, Zabalbeascoa identifies language-dependent jokes, which rely on the intrinsic natural features of the source language. They can be puns or wordplays, based on polysemy (same word, multiple meanings), homophony (same sounds but different spellings), homonymy (same sound and spelling), homography (different sounds but same spellings) and paronymy (similar sound and spelling). Humorous puns contribute to the development of the storyline and the characterisation of the protagonists. On this occasion, if source and target language are similar, they could even be translated literally; otherwise, strategies of substitution or compensation are required in order to reproduce the perlocutionary effect. Other categories identified in his classification are visual jokes, which produce a humoristic effect partially or entirely from the information conveyed visually. The analysis of these jokes is extremely relevant when dealing with audiovisual products, as the coherence between verbal and visual channel must be preserved, yet time and space constraints do not allow for expansion strategies, so in many cases a degree of humoristic loss cannot be avoided (Rollo, 2014, p.134). Ultimately, he distinguishes complex jokes, which combine two or more of the types of jokes mentioned above. Furthermore, Díaz-Cintas and Remael (2007, p. 227) add another category, that of aural jokes revolving around meta-linguistic features of speech, such as accents and intonation.

According to Zabalbeascoa (2005, p. 24) several variables need to be taken into account when translating humour. Firstly, shared humour needs to be identified, that is to say humour that can work in both languages, resorting to an almost literal translation. Translators also need to evaluate the difference in receiving audience, whether the humours is spontaneous or not, implicit or explicit. Furthermore, it is fundamental to consider the connection to the plot and the involvement in the characterisation of the protagonists, as well as the cultural differences, which might render some humoristic expressions problematic or even taboo in the target culture. In contemporary audiovisual comedy, these dynamics are deeply intertwined with the characters’ social and cultural background. Consequently, evaluating how the translators

attempt to reproduce the humoristic effect into the target text represents an object of interest for the present study. While the present study does not primarily focus on humour, this effect is contextually analysed within the broader examination of culture-bound markers in TV dialogue.

1.8 Theoretical Synthesis and Analytical Framework

Subtitling as a form of constrained translation demands that the target text continuously adapts to the spatial and temporal limitations of the screen while simultaneously operating a diamesic shift. However, this variation in dialogue from the acoustic channel to the visual and written one presents points of interest for a linguistic analysis. Specifically, as discussed in detail in the sections above, when reproducing dialogue in audiovisual products, subtitlers find themselves struggling with a text written to be performed as if it were spontaneous and consequently presenting oral features that require some form of negotiation in their Italian rendition. Nevertheless, the technical boundaries of audiovisual translation frequently oblige subtitlers to resort to strategies of partial or total reduction (as extensively documented by Díaz-Cintas and Remael) to fit into the quantitative limits of the medium, as well as omitting elements considered less relevant to the denotative and narrative meaning of the discourse. Furthermore, as translations do not simply imply a transfer from one language to another, but rather from one culture to another, problems of neutralisation and flattening of the cultural identity of the characters are common.

The present study focuses on systematically putting to the test how these linguistic and technical struggles can be overcome in a TV genre that heavily relies on the dialogic dimension: contemporary comedy. When observing the two selected TV series, *The Marvelous Mrs Maisel* (Amazon Prime Video, 2017-2023) and *Fleabag* (BBC Three / Amazon Prime Video, 2016-2019), specific linguistic features immediately strike the viewer's attention. On the one hand, the first series presents a lexically dense discourse enriched by the use of cultural markers of Yiddish; on the other hand, *Fleabag* relies on speech fragmentation (interruptions and hesitations) and a strong presence of marked speech. While these stylistic and oral markers result from preliminary observation, an empirical investigation is required in order to verify their systematic presence and evaluate how subtitlers adapt the text into the target culture and whether the expected tendency towards neutralisation and text reduction can indeed be observed. Consequently, so as to test these theoretical assumptions, the present research resorts to a corpus-based analysis through customised corpora.

More specifically, as briefly anticipated, the management of lexical reduction will be addressed on a quantitative level, through the investigation of the phenomena of lexical density

in one series and speech fragmentation in the other. Furthermore, the potential omission of discourse markers as conventional signs of orality and the neutralisation of profanities will be investigated, according to Díaz-Cintas and Remael's framework. Subsequently, the rendition of marked speech and the preservation of their effect will be investigated applying the same theoretical framework to the use of profanities and British idiomatic and slang expressions in *Fleabag*. Finally, Aixelá's exhaustive taxonomy of culture-specific items and Florin's classification of realia will be employed to analyse cultural occurrences of Yiddish realia in *The Marvelous Mrs. Maisel*, analysing their occurrences in the corpus and selected examples. These three main analytical dimensions outlined in this chapter provide the analytical framework to systematically evaluate, together with the corpus-based methodology, how fansubbers (and occasionally professional subtitlers) manage to render these linguistic features even within the boundaries of this technically constrained mode.

CHAPTER 2 - CORPUS-BASED LINGUISTICS AND TRANSLATION STUDIES: A THEORETICAL FRAMEWORK

2.1 Translation Studies: Premise

For much of the twentieth century Translation Studies revolved around the relationship between source text (ST) and target text (TT) resulting in the notion of equivalence as “formal correspondence of grammatical and syntactical structures” (Laviosa, Liu, 2021, p. 6). Consequently, the primary aim of a translation was that of maintaining as much fidelity as possible towards the core of the original work and its message. Starting from the second half of the century, a shift in orientation began to take place pushing towards a more descriptive approach to research and resulting in a target- and culture-oriented perspective where translations take on the purpose of filling the gaps between the source and target culture. Up until that time there was a tendency to ignore regularities of behaviour; on the contrary, this change in the approach aimed at producing replicable and testable data. Toury (2012, p. XI) sets the foundation for Translation Studies as a rigorous empirical discipline, based on verifiable real facts, rather than speculations or hypotheses, aiming at establishing general principles to explain and predict phenomena in the world.

From these descriptions derives a set of general principles that Toury identifies as translational norms, that is to say regularities that govern the choices made by translators within a specific temporal and contextual setting. These norms are not rigid rules, but rather patterns of use emerging from the systematic analysis of a translating community or a large body of translated texts (2012, p. 81). More specifically, Toury distinguishes between preliminary norms, concerning a general translation policy and a selection of what to translate, and operational norms, affecting the act of translation per se. Among operational norms, he introduces a further distinction between matricial norms and textual-linguistic norms. The former rule the structural organisation of the text and the distribution of linguistic material throughout it; the latter, govern the selection of linguistic material used to replace the source elements. In particular, matricial norms can draw attention to systematic tendencies of omission, condensation and manipulation in the text; whereas textual-linguistic ones to the recurrence of standard fixed collocations over creative uses, domestication over foreignising, or vice versa. This distinction is particularly productive for corpus-based research as matricial norms identify tendencies of omission and condensation in translated texts through frequency analysis, while textual-linguistic norms occur in collocational patterns recovered through concordance analysis.

This shift from meaning to usage, from conceptual to functional equivalence, inevitably requires a move towards the analysis of substantial amounts of real data in context. Baker (1993) operationalises Toury's descriptive programme, arguing that Corpus Linguistics could provide the tools needed to study translational behaviour empirically and at scale, an approach that stands at the kernel of Corpus-Based Translation Studies (CBTS). In her foundation 1993 paper, Baker proposes corpus methodologies as tools to analyse universal features of translation, that is to say features that typically occur in translated texts, regardless of the languages and cultures involved. These translation universals include explicitation, which encompasses an overall tendency for making implicit information explicit in translated texts resulting in an expansion of the target text, and simplification, reflecting a preference for transparent lexical and syntactical structures, breaking up long sentences or using a more standard vocabulary. Further tendencies involve normalisation, highlighting an inclination towards conforming to conventional grammaticality and omission of idiosyncrasies in the original text, and avoidance of repetition, by omitting or rewording repeated elements. Finally, Baker identifies a tendency to exaggerate features of the target language; and third-code operation, resulting from the confrontation between ST and TT and their convergence towards a common middle ground.

It is crucial to emphasise that these universals do not constitute fixed evidence, but rather hypotheses to be tested. In the context of subtitle translation, these tendencies interact with the spatial and temporal constraints of the format, as will be examined in the analysis of the two series in the present study.

2.2 A methodological distinction: Corpus-based vs. Corpus-driven approach

The tension between hypothesis and evidence that was mentioned above lies at the heart of a pivotal methodological distinction in Corpus Linguistics: between a corpus-based and a corpus-driven approach. On the one hand, McEnery and Hardie (2011) define corpus-based studies as typically using "corpus data in order to explore a theory or hypothesis, [...] in order to validate it, refute it or refine it" (2011, p. 6). On the other hand, Tognini-Bonelli (2001), who developed this distinction systematically, argues that in a corpus-driven approach:

the linguist uses a corpus beyond the selection of examples to support linguistic argument or to validate a theoretical statement. [...] The commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence (2001, p. 84).

In other words, a corpus-based methodology avails uses the corpus to confirm pre-formulated theories and verify their accurateness, whereas in a corpus-driven study the researcher approaches the corpus unbiased and lets the data reveal peculiarities to examine so as to establish regularities and systematic patterns in the selected texts.

The present study adopts a predominantly corpus-based approach, as the theoretical framework guides the analysis of the audiovisual products chosen. Corpora are used to test how the peculiar traits detected in the two TV shows manifest in the translated Italian subtitle version. In simple terms, the research questions of the present study are formulated as hypotheses to be tested. At the same time, following Arias-Badia's (2021) footsteps, the analysis will remain open to unexpected patterns that may emerge from corpus exploration and to the use of tools conceived as corpus-driven, such as the Corpus Pattern Analysis, yet ideal to examine pre-identified phenomena in the corpora designed for this research.

2.3 Taxonomy and Corpus Design

As not all corpora serve the same analytical purpose, it is deemed relevant to propose a taxonomy of the main types of corpus that might serve the purpose of the research presented. In the context of this study, the classification made by Federico Zanettin (2012, p. 10) will be taken into account.

He distinguishes three main categories of corpora based on the number of languages contained. The first category is monolingual comparable corpora, in which two sets of texts in the same language chosen according to the same criteria are compared through a collection of translated and domestic (non-translated) texts in the same language. This type of corpus is particularly useful to study features of translated and non-translated varieties of the same language. The second category identified is that of bilingual corpora, containing texts in two languages and highly useful to carry out contrastive analysis. Bilingual corpora can be parallel corpora (or translation corpora) in which source language texts are paired with their respective translations in the target language, allowing the identification of the translation strategies adopted in the target language. Bilingual corpora can also be comparable corpora, in which two corpora, paired on the basis of textual similar or of domain are juxtaposed.

The third category pinpointed by Zanettin is multilingual corpora in which more than two languages are involved. Like bilingual ones, multilingual corpora can be either parallel or comparable. Moreover, when bilingual and multilingual corpora are translatable in both directions, they are defined reciprocal corpora.

Finally, Zanettin distinguishes another category of corpora, which, while not created specifically for translational purposes, can nonetheless become of use for contrastive analysis. These are general reference corpora designed to be representative of a language variety and recognised as a reference point for standard use of language.

Given that this study aims at identifying the rendition strategies adopted in the process of translating from English to Italian the subtitles of two TV series and examining how the pinpointed linguistic features are rendered in the target language, a parallel corpus is the most appropriate as it allows direct alignment between source and target texts at the utterance level. Furthermore, general reference corpora, specifically the British National Corpus (BNC, tagged by CLAWS⁸) and the Italian Web Corpus (itTenTen20), will be employed throughout the present analysis. In particular, the BNC can be regarded as a reliable source not only due to its substantial size (comprising 100 million words), but also thanks to the diversity of written and spoken samples collected from a wide range of sources.

A further distinction relevant to the application of Corpus Linguistics to audiovisual translation concerns the modality of the corpus. Since audiovisual products combine verbal, acoustic and visual channels, they are de facto multi-modal products that call for a study of non-verbal language, other than the solely textual material. According to Soffritti (2019) multi-modal corpora capture the full semiotic complexity of the audiovisual product by recording and annotating multiple communication modalities, including speech, gaze, gesture and body language. These multi-modal corpora are mostly monolingual and can include annotated records of spoken communication, such as recordings, video-streaming or transcriptions of verbal content. Overall, they are designed to address not only the linguistic features of the verbal channel, but also para-verbal signs in the acoustic channel and non-verbal elements on the visual channel. Therefore, they are not restricted to the address of typical morphosyntactic linguistic features, but attention is given to intonation, pronunciation, pitch, as well as facial expressions, body movements and proxemics.

It follows that in a parallel multi-modal corpora for audiovisual translation, the elements of the source-text and their rendition in the target language need to be examined not only from a verbal perspective, but also in terms of visuals and acoustics and their coherence to the original product. While there is no denying that multimodal corpora would offer a more detailed resource of analytical study for audiovisual translation, their construction requires a level of

⁸ CLAWS stands for Constituent Likelihood Automatic Word-tagging System, a POS-tagging system for computational linguistics that converts plain text into linguistically annotated text.

specialisation and annotation that exceeds the scope of the present study. The corpus used here will, therefore, be a monomodal written corpus containing only the textual layer of the subtitles. As Pavesi (2019) demonstrates with her study, a monomodal corpus of subtitle files remains sufficient for the identification of systematic translational patterns proposed in this study. The implications of this choice are further discussed in Section 2.8.3.

2.4 Corpus Design Issues: Representativeness, size, authenticity, sampling, balance, copyright

Any corpus, regardless of its type, raises a series of design issues to address. During the course of the years, scholars have approached this question of corpus design from different angles, though recurring issues emerge across the literature. According to Kennedy, general corpus design issues are “concerned with the validity and reliability of research based on a particular corpus, including whether that corpus can serve the purposes for which it was intended” (1998, p. 60). Consequently, it follows that compilers need to identify coherent criteria to select relevant material to include in the corpus.

Among the identified issues in corpus construction, representativeness is the most widely debated one, featuring across the works of various researchers (see e.g. Tognini-Bonelli, 2001; Olohan, 2004; Peromingo et al., 2014 and Pavesi, 2019). These studies agree that a corpus should be compiled to be as representative as possible of a genre, population or category in order to be able to make generalisations based on the analysis of the represented data. Nevertheless, it can be difficult to apply objective measures that actually confirm this degree of representativeness of the texts. A strictly related issue, in fact, is that of the size of the corpus. One might assume that the bigger the corpus, the wider the chances of it being representative of the population. However, in some cases, a smaller specialised corpus might be more balanced and useful than a larger one. For this reason, authenticity and sampling are also taken into consideration as necessary conditions for representativeness. It is crucial for the selected material to be taken from genuine and authentic communication (written or spoken). Moreover, the sampling criteria used to assemble the corpora should always be made clear to the users (Tognini-Bonelli, 2001). Furthermore, a corpus can be static or dynamic, that is to say compiled of a series of texts that respectively represent language in a specific period of time, or representing language as a whole. The first case is also defined a snapshot corpus as it literally provides a snapshot of a culture and language in a selected time period; whereas with a dynamic corpus or monitor corpus, “the whole state of language can be passed before one’s eyes” (Sinclair, 1991, p. 25) as these corpora keep developing just as languages do.

Leech (2007, p. 136) identifies balance as one of the key features of a representative corpus; that is to say, a corpus that provides a range of text types wide enough to manifest patterns of usage and comparability. However, comparability and representativeness can easily be in conflict with one another. This tension will be further revisited in the following chapters in relation to the specific design choices made for this study, as well as another final issue particularly relevant to the present corpus research: that of copyright. In the case of AVT specifically, the use of commercially distributed audiovisual material raises questions about the legal accessibility of subtitle files for research purposes, which will be addressed in Chapter 3.

2.5 Corpus Pattern Analysis: Norm vs. Creativity

The analytical method adopted in this thesis is Corpus Pattern Analysis (CPA), originally developed by Patrick Hanks (2012) and applied to the study of audiovisual translation by Arias-Badia (2015, 2021). At the core of this methodological approach, there is the idea that “words are central to the activity of making meanings” (Hanks, 2012, p.1). In fact, one word does not simply have one meaning, but it changes in context.⁹

CPA is a technique of analysis that allows the researcher to capture the binary distinction between norm and creativity presented in the title of this section. It associates the normal meaning of a word with different patterns of usage (norms) and at the same time isolates irregularities (creativity or exploitation of norms). As a matter of fact, CPA is grounded in the Theory of Norms and Exploitations (TNE), also defined as a “double helix” theory of language. This theory proposes that when “making meanings”¹⁰, people can either conform to conventional patterns of ordinary usage, or exploit these norms to produce creative or humorous effects or deal with unusual topics. This distinction might erroneously be associated to the opposition between standard and non-standard uses of language. However, in Hanks’s viewpoint, a norm is systematically observable in a large corpus and, on the contrary, an exploitation deviates from this statistical regularity precisely to create contrast with the regular pattern. It is specifically the unusual character of these collocations that makes them analytically interesting. The analytical procedure applied in CPA is based on Sinclair’s (2003, p. XVI) proposal of seven-step guidelines involving the scanning and interpretation of concordances

⁹ According to Hanks, referencing the Gricean mechanism, meaning and, therefore, context are constructed around the verb. For this reason, he uses CPA for the first time to analyse different combinations and collocations around verbs. He creates an account of 680 verbs in his *Pattern Dictionary of English Verbs*.

¹⁰ Echoing the title of Hanks’s work “How People Use Words to Make Meanings” we are referring to.

and patterns (1. initiate, 2. interpret), the verification of evidence in support of the hypothesis formulated (3. consolidate, 4. report) and the refinement of the process by making adjustments if required (5. recycle, 6. result, 7. repeat). While CPA was originally developed as a predominantly corpus-driven methodology, the present study applies it in a corpus-based fashion, using it as an analytical tool to examine pre-identified phenomena in the two audiovisual products at the core of the research.

This type of analysis is strictly connected to Laviosa's (1994) concept of neutralisation in translation. She defines it as a tendency to conform to the normative use of words, rather than searching for an equivalent creative and ambiguous rendering in the target language. Arias-Badia (2015) puts the verification of this concept into practice in her work (2020) on Spanish subtitling, identifying how in most cases the creative speech of characters is easily neutralised producing a more neutral target text. This tendency could be attributed to the constrained nature of audiovisual translation and, more specifically, subtitling. In fact, the spatial and temporal limitations of the format (discussed in Chapter 1) can produce phenomena of condensation and/or neutralisation, leaving aside the creative exploitation of the language.

The concept of neutralising speech of a creative character can be considered as particularly productive for the present analysis. As explained in more detail in Chapter 3, the TV series under analysis are marked by linguistic features such as speech fragmentation, lexical density and the presence of realia. If the Italian subtitles systematically undergo a neutralisation of these characteristics, the target audience would receive a distorted translation of a completely different version of the characters. With this corpus analysis, it will be possible to assess the degree of systematic neutralisation operated in the series.

The notion of lexical density mentioned above involves the ratio of content words to total words in the text. In the context of comedy subtitling, lexical density is relevant as a measure of the complexity of the source text, as well as a symptom of translational shifts and strategies. Usually, this parameter shows that written texts tend to exhibit a higher degree of complexity because of a more recurrent use of nominal structures in comparison with spoken language. In order to verify this distinction, Arias-Badia applies the type-token ratio (TTR) formula, that is to say the number of word types (lemma) divided by the total number of tokens, as a measure for vocabulary richness. Therefore, a lower TTR in the target text compared to the source text is evidence of a lexical simplification. However, as it will be demonstrated in the following chapter, one of the two products analysed in this project displays a lexical density closer to written prose rather than spontaneous speech, as a result of the fast-talking style of

one of the main characters. Furthermore, the choice of using fansubs rather than professional subtitles plays a pivotal role in this analysis as well.

2.6 Computational Tools for Linguistic Analysis: Sketch Engine

Sketch Engine is a corpus tool launched in 2003-2004 by Adam Kilgarriff and Pavel Rychlý and described in its technical presentation by Kilgarriff et al. (2014). The company's tagline "Corpora for all" reflects its core philosophy: Sketch Engine refers to both the software and the web services constituting a platform of a wide range of corpus analysis tools accessible to researchers, translators, language teachers, terminologists and lexicographers. It is the computational tool adopted to compile, query and analyse the parallel corpora designed for the purposes of the present research.

The platform supports different types of corpora relevant to this study. Firstly, preloaded corpora, managed directly by the Sketch Engine team, include large general language web corpora (usually comprising at least 50 million words), kept up to date, for about sixty languages, as well as reference corpora, such as the British National Corpus (BNC) and the Brown Corpus, used as a benchmark for keyword analysis and collocational norms in standard language. Thirdly, Sketch Engine provides a platform for parallel corpora, the primary configuration used in this study. Through parallel concordancing, that is to say the alignment of texts that are translations of each other, the researcher is able to systematically visualise how one term in the ST is rendered in its translation in the TT. Lastly, user corpora, also called do-it-yourself (DIY) corpora, are corpora uploaded and compiled by researchers for specific analysis purposes. As already mentioned, such is the case of the present study: two parallel DIY corpora.

Sketch Engine allows for a multi-layered interrogation of the data through several key features. When uploading the raw texts to compile and process a corpus on Sketch Engine, the software automatically initiates a three-stage pipeline of tokenisation, in which the texts are split into single units, lemmatisation, to reduce inflected forms of each word to their base "dictionary" version, and POS-tagging to assign grammatical labels to each word. The product of this process is an automatically annotated version of the raw texts forming the corpus.

2.6.1 Frequency Lists and Keyword Analysis

The most basic quantitative procedure of corpus analysis is the creation of frequency lists, which allow the researcher to identify "how often each different word-form occurs in a text"

(Sinclair, 1991, p. 30). In Sketch Engine, frequency lists can be generated for each corpus and can be useful, for the purposes of this study, to provide an overview of the most characteristic vocabulary for each series, making it possible to identify even before running any specific query, possible linguistic points of interest in each corpus.

Keyword analysis can be perceived as an extension of frequency listing, as a keyword is a word that occurs frequently in a specific corpus when compared to a reference corpus. Baker (2020, p.130) proposes the use of a statistical method called Log-Likelihood Test (LLT) in order to compare the frequency of terms in the DIY corpora to a larger reference corpus (for instance, the BNC). The generated file presents in line each item with its frequency in the DIY corpus, followed by how many times it occurs in the selected reference corpus, the relative frequency in the DIY corpus, the relative frequency in the reference corpus and, finally, the LLT score that measures the difference in frequency between the two corpora. The higher is the score, the more characteristic is the term in the DIY corpus compared to its reference one, with the results obtained appearing in descending order.

Using this test, slang terms and profane language occurring with unusually high frequency in the DIY corpora in comparison to the general use of language, can be easily identified. Therefore, a keyword analysis helps identifying the vocabulary that is distinctive of the corpus under study. For the purposes of this research, it can be applied to both English and Italian corpora so as to provide a starting point for the analysis of the most distinctive lexical items.

2.6.2 Concordance, Corpus Query Language

The foundational tool of Corpus Analysis is the concordance, that is to say a “list of all the occurrences of a word, phrase or sequence of word/tag combinations in a corpus with each one occurring with a few words of co-text” (Baker, 2020, p. 135) in a format known as keyword in context (KWIC). More specifically, each node appears at the centre of each concordance line with up to five words on either side of the search item, allowing the researcher to observe the typical contexts in which the term occurs. Sketch Engine is a fourth generation concordancer¹¹ modelled on the Google search form, making simple searches case-insensitive and allowing queries by word form or lemma.

¹¹ McEnery & Hardie (2011) provide an exhaustive description of the evolution of concordancers throughout the years. Sketch Engine is included in the fourth-generation category as it shifts the focus to large-scale web-based research and the processes of POS-tagging, lemmatisation and alignment are automatic.

For more complex queries, Sketch Engine adopts Corpus Query Language (CQL), based on the formalism developed in the 1990s at the University of Stuttgart. CQL allows users to search for more specific queries based on lexical, morphological or syntactical criteria. Once a concordance has been generated, Sketch Engine makes it possible to sort it and filter it in multiple ways; for instance, in alphabetical order on the words to the left or to the right of the search term. This type of filtering can be useful to identify collocational patterns if a node systematically appears with the same word or phrase to its left or right. With the term collocation McEnery & Hardie (2011, p.123) refer to “a wide range of different co-occurrences patterns that may be extracted from a corpus”, also referred to as n-grams, multi-word units, clusters or lexical bundles.

2.6.3 Collocations and Word Sketches

As outlined above, in Corpus Linguistics the tendency of words to occur with specific other words is referred to as collocation. As a matter of fact, words do not carry meaning on their own, but rather in association with other words or phrases. This is called co-occurrence and the study of this phenomenon reveals the typical semantic environments of source and target language.

Baker (2020) proposes the adoption of another statistical measure called Mutual Information (MI), related to the Log-Likelihood test linked to keyword analysis. The difference between the two is that the Log-Likelihood Test measures the relevance of a single item, whereas MI captures the strength of the association between two items. Applying this analysis might be significant to identify idiomatic slang phrases and recurring patterns in the use of profane expressions. This distinction demonstrates how these two measures can be used in complementary ways in this research.

The systematic study of collocational patterns of a specific linguistic item is made possible in Sketch Engine specifically through the function that gives its name to the platform: the word sketch. As presented by Kilgariff et al., a word sketch is “a one-page summary of a word’s grammatical and collocational behaviour” (2014), which provides an overview of the recurring patterns of the word, the verbs it collocates with and the adjectives that modify it. Applying these features of corpora analysis to the parallel corpora designed for the present study will make it possible to identify systematic translational patterns in the Italian subtitles that would not emerge from the examination of individual translations on their own, as it will be shown in the following chapters.

2.7 Using Corpus Analysis in Audiovisual Translation

Pavesi provides an exhaustive list of reasons to adopt a Corpus Analysis for Audiovisual Translation. Corpus Based Translation Studies (CBTS)¹² “reveal systematic patterns, distinctive features, norms and universals in translated texts by relying on extensive authentic data” (2019, p. 315) that, as mentioned before, would not be as evident through individual manual examination. Therefore, corpora become useful instruments to exploit in order to improve the quality of the translations produced.

First and foremost, the empirical nature of the field not only allows the researcher to go beyond single intuitions, but through the repetitions of the same patterns across texts, culture- and ideology-bound behaviours in translation can be identified, as well as hidden and opaque meanings. This systematic exploration using corpora takes place through tools such as frequency counts, to discover patterns of translation and monolingual and bilingual concordances for a more detailed analysis. Designing corpora for audiovisual translation purposes, which are usually small DIY corpora, presents the same building issues examined more broadly through CBTS, including representativeness, size, comparability, sampling. Dealing with spoken corpora, AVT corpus design has to pay even more attention to issues of transcription, corpus alignment and annotation. When dealing with subtitles, copyrights become a serious issue to tackle due to the difficulty and time-demanding process of the acquisition of rights to the official subtitle sources. As seen in the present study, researchers often find themselves resorting to open access fansubs as an alternative source.

This systematic type of research has led to the documentation of several recurring strategies audiovisual translators are more inclined to go for, such as standardisation, interference, simplification, omission, equivalence, substitution and creative reformulation. For instance, these tendencies were a result obtained focusing especially on typical features of dialogue, such as discourse markers, which tend towards strategies of standardisation and neutralisation, together with marked language. Taking a deeper look into these observed trends, standardisation refers to the suppression of marked features or non-standard expressions; interference to the presence in the target language of features typical of the source language; omission, simplification and lexical reduction are required by space constraints. All of these tendencies are perfectly consistent with the audiovisual translation strategies examined in detail in Chapter 1, through the work of Díaz-Cintas & Remael (2007) and Aixelà (1996).

¹² The term Corpus-Based Translation Studies was coined by Baker in 1993.

These specific tendencies, together with the constraints of the subtitle format, constitute a set of genre-specific patterns at the core of the present research. Through the systematic analysis of these choices across the parallel corpora of both series, it is possible to assess the degree to which these tendencies manifest in the Italian subtitles adopted here.

2.8 Corpus Design and Selection Criteria

When conducting a corpus-based research it is deemed pivotal to dedicate an exhaustive section of this thesis to motivating and describing the construction of the bilingual monomodal parallel corpora on which this research is based. As stated by Sinclair, “the beginning of any corpus study is the creation of the corpus itself” (1991, p. 13).

The research work presented in this thesis rests on two English to Italian parallel corpora, one for each of the two series to analyse. Each corpus is composed of the complete run of episodes from the two contemporary English-language comedy series: *The Marvelous Mrs. Maisel* and *Fleabag*, together with their respective Italian subtitle translation made by fan subbers (non-professionals). As discussed in Chapter 3, the two series were selected so as to represent two distinct sub-genres of contemporary streaming English-language comedy, each defined by a different set of linguistic features that pose a specific challenge for subtitling.

2.8.1 Corpus Size

The choice of creating a corpus including the entire catalogue of episodes is consistent with the principle of representativeness discussed in the chapter above: a complete corpus of available material allows a systematic study of patterns of repetition.

The complete episodic run of *The Marvelous Mrs. Maisel* comprises 5 seasons for a total of 43 episodes (43-76 minutes), whereas *Fleabag* is made up of 2 seasons for a total of 12 episodes (22-27 minutes). The asymmetry in length of the two series becomes analytically relevant as well. The first series accumulates a corpus across five years that reflects the evolution of the series over time; by contrast, *Fleabag* provides a very compact corpus, drawing on its linguistically fragmented nature.

SERIES	SEASONS	EPISODES	EN TOKENS	IT TOKENS
The Marvelous Mrs. Maisel	5	43	466,004	387,181

Fleabag	2	12	42,437	36,214
---------	---	----	--------	--------

Episodes and token count in both TV series.

2.8.2 Data Sources and Fan-Produced Subtitles

The subtitle files used herein were downloaded from the website OpenSubtitles¹³, the biggest multi-language (more than 30 languages) subtitle database. Both the English source subtitles and the Italian target subtitles for the creation of the two corpora were obtained from this source. These files were downloaded in SubRip Text format (.srt), including timestamps and text strings. It should be noted that the subtitles available on this platform are fan-produced. The decision to adopt fan-generated subtitles (presented in the first chapter of this work) as corpus material, rather than professional subtitles made by official platforms such as Amazon Prime or BBC, lies in their unfiltered linguistic representation.

While an official translation is more bound to be subject to that neutralisation process mentioned in Chapter 2 in order to comply with commercial or branding guidelines, fansubbers tend to adopt more creative strategies of translation, especially when rendering taboo or humorous language. Nevertheless, using Venuti's (1995) renowned opposition, fansubs can be considered typical examples of foreignization, whereas official subtitles prefer to adopt a normalisation strategy and produce a more domesticated product. Consequently, for a study focused on the translation of profanity and slang, fan-produced subtitles provide a fertile ground for observations, as they tend to be more inclined to preserve cultural references, slang expressions and linguistic idiosyncrasies in general.

Furthermore, on practical grounds, official subtitle files for streaming productions are not publicly accessible for redistribution and their use would raise copyright issues (Chapter 2). In conclusion, the fansubs constitute the primary material of the corpora created. However, if deemed interesting, individual extracts of particular instances of translation in the official versions will be consulted through the streaming platforms and manually transcribed here, comparing the solutions proposed by the two versions and, if relevant, proposing alternative options.

¹³ (<https://www.opensubtitles.org/>)

2.8.3 Pre-processing, Compiling and Metadata Structure

Generally SRT subtitle files contain, other than dialogue text, a series of non-linguistic elements, such as timestamps and accessibility captions for hearing-impaired, which need to be removed as they could interfere with the corpus analysis. The latter could include the description of sounds in brackets (e.g. *[upbeat music]*) or speaker identifiers for who is speaking. Consequently, a pre-processing procedure is required in order to clean up the subtitles of these non-linguistic labels. Therefore, before being uploaded into Sketch Engine, the raw files downloaded from OpenSubtitles underwent a cleaning process in *Subtitle Edit*¹⁴ to eliminate most of these elements.

After the removal of non-dialogue elements, the cleaned SRT files were exported in plain text format and uploaded on Sketch Engine to assemble the two parallel corpora, each comprising source text material in English and target text material in Italian. As already discussed in Chapter 2, in a parallel corpus English and Italian files are stored separately, enabling monolingual queries, as well as bilingual comparative queries. The decision to create two independent corpora sharing the same scheme, rather than a single corpus with subcorpora for each series, allows for an easier analysis: in this way, each series can be examined for its own distinctive linguistic features, also taking into consideration the different length of the two objects of study.

Once uploaded to Sketch Engine, each file was processed through the platform and automatically underwent a procedure of tokenisation, lemmatisation and POS-tagging, integrated into the platform (see Chapter 2 for a detailed description of these procedures). The product of this automatic linguistic annotation enables a series of corpus queries useful to identify discourse markers, fragmentary structures and linguistic patterns.

Prior to the upload, all the files were named following the same scheme: series title, season and episode number and language (e.g. *Maisel s01e01 EN*, *Maisel s01e01 IT*, *Fleabag S01E01 EN*, *Fleabag S01E01 IT*, etc.) to enable Sketch Engine to allow these queries and filter them by season and episode, as well as aligning source and target files at episode level for parallel concordance queries.

2.8.4 Limitations of the Corpus Design

As previously mentioned, designing a corpus always implies some limitations. In this instance, the use of fan-produced subtitles downloaded from OpenSubtitles bypasses copyright issues,

¹⁴ Subtitle Edit is an open-source subtitle editor. It can be used to synchronise subtitles to a video, make new subtitles from scratch, convert different formats, and much more. (<https://www.nikse.dk/subtitleedit>)

but might introduce variables related to the competence and the stylistic preferences of a non-professional translator. As a response to this issue, a targeted consultation of the official subtitles can compensate this limitation, yet it is relevant to notice the implications of such a choice.

A further limitation concerns the alignment of source and target subtitle files. Fan-produced subtitles occasionally present misalignments between the English and Italian segments, which may affect the reliability of parallel concordance results. All instances selected for qualitative analysis were therefore manually verified against the original source material before classification. The issue of multimodality was also mentioned before. Even though the present study deals with multimodal products, monomodal corpora are used for this analysis, therefore concentrating on their textual level. These limitations are characteristic of corpus-based audiovisual translation, as acknowledged in Pavesi's (2019) study of the field.

2.9 Data, Analytical Tools and Methodological Summary

Following the description of the corpus design and compilation criteria presented in the previous sections, this final section of the chapter is dedicated to the exact investigative framework adopted and the sequence of operational steps taken to analyse the linguistic phenomena, to make sure that the corpus tools serve as empirical instruments for the following study. The operational analysis is initiated through a quantitative macro-evaluation of the corpus data using keyword analysis. These tools serve as filters to isolate the systematic lexical patterns across the source and target text that will constitute the starting point of the entire examination.

Following this initial statistical filter, wordlists are generated for each of the four corpora of data to calculate the number of tokens and types of words. By applying the Type-Token Ratio (TTR) formula, this step serves the purpose of confirming or refuting the hypothesis of text reduction and consequent condensation or reformulation strategies. Once these reduction rates are generated for both TV series, potential causes for token reduction can be examined according to the AVT literature presented in Chapter 1, primarily focusing on the omission of discourse markers and marked language.

Concordance tools will be used to highlight phenomena of high lexical density in *The Marvelous Mrs. Maisel* by detecting continuous sequences of words or multi-word units, further analysed through parallel concordance queries to track how they are rendered in the target texts in Italian, with the assistance of reference corpora to consider alternative solutions that might prevent potential translational losses. Moreover, wordlist tools allow the investigation of non-

alphabetic characters, that is to say of punctuation markers, which in AVT are treated according to their own specific grammar conventions, as mentioned in Section 1.4.1.. Corpus Query Language (CQL) strings, primarily of dashes (-) and suspension points (...), can be useful indicators of speech phenomena such as hesitations, repetitions and interruptions, producing the so-called fragmented speech in *Fleabag*. Furthermore, this step allows the researcher to verify whether these conversational features are preserved or neutralised in the target text and what the implications of such choices are.

The series just mentioned further shows peculiar linguistic features in register through the systematic use of swearwords and idiomatic expressions. Profane words can be analysed through a query at lemma level within the parallel concordance tool, a device that assesses whether peculiar utterances of taboo language are maintained or adapted in the target culture. Idiomatic or slang expressions as a linguistic category, on the contrary, cannot be put through automated querying and will require a manual selection of productive samples from the source material. This step of the research adopts a corpus-assisted approach, as the parallel concordance tool will then be employed to identify and evaluate the Italian renditions. In addition, where fan-produced solutions appear inadequate, a contrastive analysis with the rendition made by professional subtitlers on official streaming platforms can be adopted for wider investigation of translational challenges and evaluation of contrasting tendencies.

Finally, as the research deals with strongly culturally-marked case studies, a focus will be put on Yiddish realia using complex CQL queries for each lemma. The extracted lines of fansubs will then be confronted with manual transcriptions of the official renditions, establishing patterns towards domesticating or foreignising tendencies. This methodological framework adapts theoretical assumptions and intuitions to an investigative approach, primarily based on quantitative analysis, yet open for follow-up in-depth qualitative examination.

CHAPTER 3 - TWO CASES OF STUDY: AN ANALYSIS OF *THE MARVELOUS MRS. MAISEL* AND *FLEABAG*

The present analysis investigates the translation strategies adopted in the Italian subtitles of two TV series: *The Marvelous Mrs. Maisel* (Amazon Prime Video, 2017-2023) and *Fleabag* (BBC Three / Amazon Prime Video, 2016-2019). Both series share a number of features that make this joint analysis and comparison productive, yet present distinct comic registers and cultural and linguistic traditions that pose different sets of challenges for the Italian subtitler.

Both works are written from a female perspective¹⁵ and were produced and distributed within the same decade (2015-2025), the era of mainstream streaming platforms. However, they are set in different sociocultural contexts: Mrs. Maisel's story is set in the late 1950s in the Jewish community of New York with its cultural and religious specificities, whereas *Fleabag*'s life story unfolds in contemporary London. Both series use humour as a vehicle for exploring sensitive themes such as female identity, social expectations, failed relationships, grief and professional marginalisation, painting the two comedies with a darker undertone.

The analysis presented here draws mainly on quantitative corpus data, supported by the examination of parallel concordances with particular attention to four phenomena: lexical density and the rendering of Yiddish realia in *The Marvelous Mrs Maisel*, and speech fragmentation and the management of slang and profanities in *Fleabag*.

3.1 Starting point: Keyword Analysis

The starting point of the present analysis that will lead the way for all consequent and fuller examination lies in the compilation of frequency lists for each TV series corpus, both in the source language (English) and target language (Italian), for both series under review. The tool at use in this case is called wordlist with a basic query of all words, which generates a frequency list of all tokens in the corpus in order to identify potentially salient terms within its vocabulary. This tool was used primarily to generate the number of tokens necessary to apply the type-token ratio formula discussed in Section 3.2.

In the present case, a further analytical step was carried out before drawing the conclusions for each series: a keyword analysis. As explained in detail in Chapter 2, the keyword analysis uses the Log-Likelihood Test as the default statistical measure for compiling

¹⁵ Amy Sherman Palladino and Phoebe Waller Bridge, respectively, are the sole authorial voice behind the TV series.

a list of words occurring with frequency in comparison to a selected reference corpus. In this case, the reference corpora used for the English language is the British National Corpus (BNC), and more specifically, its subcorpora Spoken, a smaller collection of around 10 million words in the bigger corpus of informal conversations. The reason for this choice lies behind the idea that subtitles simulate spoken discourse and using a reference spoken corpus is methodologically more coherent than using a general written language corpus. However, for the Italian language a corresponding spoken corpus available on Sketch Engine could not be retrieved, therefore the Italian Web 2020 (itTenTen20) was picked as it is the biggest collection of texts from the Internet available on the platform.

The *Marvelous Mrs. Maisel*'s keyword analysis brings to the surface a series of patterns. Apart from the predictable anthroponyms (*Midge, Abe, Maisel, Joel, Susie, Ethan, Weissman, Moishe*), stand-up comedy related terms such as *comedian* (LLT: score 56)¹⁶, *stand-up* (71), *onstage* (71), *gig* (69), *comic* (36), *theater* (114) confirm the series' milieu, together with Americanisms, *gonna* (2131), *favorite* (110), *color* (84), *honor* (97), *neighbor* (37), *theater* (114). The high frequency of Yiddish realia is extremely interesting for the purposes of this analysis with occurrences such as *rabbi* (95), *yom kippur* (48), *Catskill* (74), *gaslight* (117) and these will be examined in detail in Section 3.6. Fragmentation markers *i-i* (256), *uh* (184), *mm-hmm* (172), *uh-huh* (101), *y-you* (35) present high rates of frequency, supposedly as a consequence of the characteristic fast-talking of the characters. Profanities are also quite recurrent, with occurrences of *goddamn* (200), *asshole* (109), *ass* (65), *fuckin* (48), showing signs of a low and colloquial register.

Overall, the Italian corpus preserves the same patterns: character's names and surnames are among the most present occurrences. Yiddish realia hold their safe spot as they are either preserved or adapted to the target language, *rabbino* (45), *yom kippur* (47), *mitzvah* (31), *catskill* (74), as well as stand-up comedy related terms, such as *cabarettista* (70), *tournee* (44), *broadway* (32). Profanities are preserved and rendered with several alternatives *fanculo* (82), *cazzo* (58), *vaffanculo* (45), *merda* (43), *stronza* (35), *stronzo* (34), *fottere* (32), *tetta* (32).

Similar patterns resulted from *Fleabag*'s keyword analysis. Antroponyms get the highest score of frequency with *Claire* (77), *Belinda* (70), *Hillary* (109), *Jake* (105), *Priest* (68),

¹⁶ The number next to each term indicates the Log-Likelihood Test score, not the number of total frequencies of the word. The LLT confronts the frequency in the DIY corpus with the reference corpus chosen, as better explained in Section 2.6.1.

Sylvia (62), *Patricia* (58)¹⁷. Fragmentation markers are much more present here, compared to the other show under examination, *i-i* (401), *a-a* (127), *just-just* (118), *well-well* (47), *it-it* (48), *the-the* (48), *uh* (468), *um* (95), *mm-hmm* (307): all these occurrences demonstrate hesitation and frequent repetition of the same words showing signs of speech fragmentation as a systematic feature. Given the smaller size of the *Fleabag* corpus, profanities are even more recurring in the series: *fuck* (68), *slut* (162), *vagina* (89), *masturbate* (83), *asshole* (61) with higher rates of recurrence. Interesting instances of wordplay also appear repeatedly, such as *sexhibition* (236), *weaky* (48). Remarkably, in this corpus both British slang and realia *café* (153), *chatty* (56) and Americanism are highly present, *awesome* (71), *color*, *favorite*, *honor*, *theatre*. This coexistence indicates an interference of American orthography in a British TV series, which might have been either an artefact due to fansubs produced by American speakers, or reflect the influence of American English in series produced in the streaming era by Amazon Prime Video.

On the whole, this Italian corpus as well preserves the same patterns of frequency with anthroponyms at the top of the list, followed by profanities with their several Italian equivalents: *fottiti* (215), *puttana* (105), *cazzo* (101), *vaffanculo* (94), *stronza* (82), *fottutamente* (70), *scopare* (69), *fottere* (52), *merda* (45), *stronzo* (44), *vagina* (46). The same wordplay and neologism appear as their Italian equivalents: *sesposizione* (194) and *pappamolle* (82).

3.2 Wordlist and token count

As briefly mentioned above, a basic tool used to set the condition of the present research is the Wordlist option on Sketch Engine. For each of the four corpora (both English and Italian), a frequency list of all the words was generated, with the relative number of *tokens* (total frequency) and *types* of words and the type-token ratio (TTR) formula mentioned in chapter 2.5 was applied.

$$TTR = \frac{items}{total\ frequency}$$

The results obtained can be summarised in the following chart:

¹⁷ The protagonist's proper name is never mentioned in the series, she is just indirectly known by the audience as *Fleabag*. According to the definition listed in the Oxford Advanced Learner's Dictionary the noun *fleabag* refers to "a person who looks poor and does not take care of their appearance".

TV SHOW	LANGUAGE	TOKENS	TYPES	TTR
<i>The Marvelous Mrs. Maisel</i>	EN	466,004	15,320	0.0329
<i>The Marvelous Mrs. Maisel</i>	IT	387,181	21,089	0.0545
<i>Fleabag</i>	EN	42,437	2,898	0.0683
<i>Fleabag</i>	IT	36,214	4,180	0.1154

Token count and TTR: English and Italian corpora compared.

This chart displays that in both cases the TTR is higher in Italian in comparison to English. This result is in contrast with the simplification hypothesis expected according to the translation universals introduced in Chapter 2. The reason for this unexpected result might lie in the morphology of the Italian language: as a matter of fact, in comparison with English, Italian is richer in morphology with its various verb forms and adjective agreement in gender and number. As a result, given the same semantic content, an Italian text will automatically produce more types than its English equivalent due to the inflectional morphology generating different forms. Therefore, calculating the TTR might be more reliable when dealing with texts in the same language, rather than translations. However, what remains of interest for the purposes of this analysis is the reduction in tokens from English to Italian, a phenomenon that could suggest a condensation tendency. As shown in the following table, both series display a systematic reduction in token count from source (EN) to target text (IT).

TV SERIES	ST TOKENS	TT TOKENS	REDUCTION	%
<i>The Marvelous Mrs. Maisel</i>	466,004	387,181	78,823	16.9%
<i>Fleabag</i>	42,437	36,214	6,223	14.7%

Token reduction percentages from ST to TT.

The Marvelous Mrs. Maisel loses about 79,000 tokens, whereas *Fleabag* around 6,000. The percentage of condensation is very similar showing a systematic tendency in subtitling. This pattern of reduction of about 15-17% of tokens in the Italian versions of both TV series constitutes quantitative evidence of that condensation tendency discussed in Chapter 2. This

result appears coherent with the simplification theory highlighted by Baker (1993) and the condensation patterns documented in AVT research by Pavesi (2019).

The Marvelous Mrs. Maisel showing a slightly higher reduction rate in comparison with *Fleabag* (16.9% against 14.7%) supports the hypothesis of higher lexical density of the original text: a more dense dialogue requires a stronger condensation work from the translator.

Drawing on the literature discussed in the previous chapters, potential causes of this token reduction can be identified in the omission of discourse markers (Chaume, 2004a; Pavesi, 2019), syntactic condensation (Díaz-Cintas & Ramael, 2007) and the omission of marked vocabulary such as slang and profanities (Arias-Badia, 2015), all phenomena that will be discussed in the course of the present study. Throughout the analysis, the OpenSubtitles parallel corpus, as well as the BNC reference corpus mentioned above, have been consulted to verify translation alternatives and typical collocational behaviour of selected occurrences both in English and Italian. This informal application of the collocational analysis discussed in Chapter 2, provides a practical basis for evaluating the translation choices made by the subtitlers and presenting alternative suggestions.

3.2.1 *The Marvelous Mrs Maisel*: Fast-Talking and Lexical Density

Fast-talking generates a series of extremely long sequences of functional words, which result in what can be defined lexical density, a concept discussed in Chapter 1. Using the formula $[tag="NN"|tag="NNS"|tag="JJ"|tag="VB"]\{4,\}$ in the concordance tool, the research shows occurrences in the corpus of sequences of at least four (4) consecutive words belonging to the grammatical categories of singular nouns (NN), plural nouns (NNs), adjectives (JJ) and basic form verbs (VB). Consecutive means that these words are not separated by any article, preposition, pronoun or conjunction.

With this formula, 100 occurrences are shown in the corpus; switching to a parallel concordance it is possible to check how they were rendered into Italian, if preserved.

SOURCE (EN)	TARGET (IT)
“She was spinning webs around me, man, she was like a <i>evil web-spinning spider death sorceress</i> .” (S01E03, 04:35)	“Mi stava tessendo una ragnatela attorno, sembrava una <i>malefica strega</i> ragno che <i>tesse ragnatele</i> .”

“Now, you could be an original, but you are fucking it all up with <i>this cockamamie alternate universe party bullshit</i>.” (S01E06, 39:23)	“Puoi anche essere originale, ma mandi tutto a puttane con <i>questo ridicolo universo alternativo di feste di merda</i>.”
“I mean, was it really fair to lure Joel back with the promise of <i>coherent conversation and unlimited clean jockey shorts</i>?” (S01E08, 55:38)	“E' stato giusto riprendersi Joel con la promessa di una <i>conversazione coerente e boxer puliti per sempre</i>?”
“I'm gonna roll! - Then roll, <i>you stupid old expositional fake cowboy guy</i>.” (S03E04, 03:06)	“Li sto lanciando! - Allora lanciali, <i>finto cowboy stupido e sbruffone</i>!”
“He's got a six-point health care plan that includes everything from building new hospitals to spiffier paper gowns and turning the heat on so you don't freeze your tush off in the examining room.” (S04E07, 35:26)	“Un piano sanitario in sei punti che include nuovi ospedali, camici più raffinati e ambulatori più caldi per non gelarvi le chiappe.”
“No, but you created the environment: the dolls, the books about being a mommy, the quarter-scale toy kitchen you rigged up for me with the gas broiler and the little fridge that plugged into a working socket.” (S05E02, 04:49)	“No, ma mi avete circondata di bambole, libri sull'essere mamma, la cucina in miniatura che avevo, con la griglia e il frigo con una presa vera.”

Selected examples of lexical condensation strategies in *The Marvelous Mrs. Maisel*.

In the first example, the dense sequence of nouns and adjectives *evil web-spinning spider death sorceress* is rendered through a lexical condensation: the word *death* disappears and the nominal phrase is rendered through a relative clause (*che tesse ragnatele*). Following Díaz-Cintas and Remael's (2007) classification of translation strategies, this can be considered a case of reformulation that reduces the lexical density, but preserves the general meaning of the source text.

The second instance is of particular interest, as at first glance the numerical lexical density seems to be preserved. However, in this case the loss is not on a quantitative level, but more on a qualitative one. The culturally marked American slang term *cockamamie* meaning

“silly, not to be believed” (Oxford Learner's Dictionaries, n.d.) is rendered with a reformulation (Díaz-Cintas and Remael, 2007) through the more neutral *ridicolo*. Despite the evident cultural loss, when trying to look for better options for this rendition on OpenSubtitles used as a parallel reference corpus, the alternatives given for *cockamamie*, other than *ridicolo*, are *robaccia*, *stupido*, *assurdo*, *pagliacciata*. It is one of those cases, in which a direct equivalent does not exist in Italian; hence, some degree of translational loss is inevitable.

Following with the third example, another reformulation (Díaz-Cintas and Remael, 2007) is adopted with *unlimited* rendered as *per sempre* which implies a subtle shift in meaning from infinite supply in quantity, to infinite in time. Moreover, the specification *jockey shorts* is translated as a generic *boxer*, reducing the specificity of the item.

In the fourth selected occurrence of lexically dense phrases, the original adjective and noun sequence *you stupid old expositional fake cowboy guy* is condensed in its rendition into Italian as *finto cowboy stupido e sbruffone*. The adjective *old* is omitted, as well as *guy* at the end, making the overall sentence sound less lexically rich and reducing the fast-talking effect of the protagonist's speech.

In the fifth example, a long and elaborated list is systematically condensed causing a loss of the comical effect of the original utterance. From 35 tokens in the original version, it is reduced to 22 tokens in Italian making it the most significant reduction among the selected examples presented here. Furthermore, this long sentence contains a series of marked language instances, including *spiffier* and *tush* (the latter will be analysed in detail in Section 3.6). The American informal adjective *spiffy* means “attractive and fashionable” (Oxford Learner's Dictionaries, n.d.) and is neutralised here with a generic *raffinato*. Among the occurrences shown in the reference corpus OpenSubtitles consulted for further translation alternatives, an interesting choice presented might have been *in tiro*, which uses an equivalent Italian idiomatic expression to at least try to preserve the marked language of the original.

Finally, the last example is another case of systematic condensation from 33 tokens in English to 24 in Italian. In this instance, different strategies of translation are combined: terms and expressions such as *quarter-scale*, *created the environment*, *rigged up*, *gas broiler*, *plugged into* are all neutralised and reformulated into a condensed construction that preserves the general meaning but reduces the detailed description minimising the humorous effect of the fast-paced utterance. Overall, the source text is simplified into a more generic target text.

In short, the dominant pattern emerging from the examination of the selected examples is that the Italian translations tend to adopt condensation and reformulation strategies, primarily on specifying lexical constructions. As a consequence, the lexical density of the original voice

of the show is systematically reduced as confirmed by the reduction in tokens of the Italian corpus in comparison to the original one, as discussed in the section above. It is evident that the exceptional lexical density of *The Marvelous Mrs. Maisel* places a greater condensation pressure on the translator than would be expected in audiovisual products whose dialogue reproduces the lexical density norms of spontaneous spoken language. This finding demonstrates that this condensation pressure varies across AVT depending on the linguistic features of the product, rather than simply on the standard technical conventions; moreover, it inevitably affects not only the denotative content of the utterances, but also their humorous register and the distinctiveness of the character's voice, supporting Arias-Badia's (2015) framework of character voice neutralisation.

3.2.2 Frequency Lists: an Analysis of Punctuation Marks

Wordlists can be useful to provide data for TTR and token reduction, yet, differently from the keyword lists analysed in the previous section, they include in their examination punctuation marks as well. In particular, the generated lists show that other than commas and dots, dashes (-) and suspension points (...) rank among the top places in the file, right after the standard dots and commas. These two markers show interesting patterns of interruption and fragmentation of the text.

More specifically, the following chart shows the obtained results.

MARKER	TV SERIES	EN	IT	VARIATION
...	<i>Mrs. Maisel</i>	2753	2251	-18% in IT
...	<i>Fleabag</i>	679	788	+16% in IT
-	<i>Mrs. Maisel</i>	15503	13900	-10% in IT
-	<i>Fleabag</i>	1058	1505	+42% in IT

Frequency of fragmentation markers (-) and (...) across the four corpora.

This comparative table clearly illustrates opposite tendencies in the two series. In *The Marvelous Mrs. Maisel* both suspension points (...) and dashes (–) decrease in Italian demonstrating a tendency for condensation and simplification. In *Fleabag* both increase in Italian, reproducing the speech fragmentation of the original text.

It might be assumed that the translator uses them even when they are not present in the original, perhaps for other purposes related to subtitles' conventions (Section 1.4.1), such as the use of dashes to indicate a change in speaker. To verify this hypothesis a *[word="-"]* CQL

query in advanced concordance can be used to search for the dash and skim through some random samples in the corpus. The results obtained from this analysis show that dashes are mostly used to identify the change of speaker, both in English and Italian, whereas suspension points [*word="\.\.\."*] are employed to indicate interruption. These data, therefore, show that dashes cannot be considered signs of speech fragmentation useful to confirm this assumption. Suspension points, however, remain reliable indicators of fragmentation and will be examined in greater detail through the following parallel concordance analysis in the *Fleabag* English corpus.

3.2.3 *Fleabag*: Speech Fragmentation and Interrupted Sentences

A random selection of 250 occurrences of (...) in the corpus shows that the translator recognises syntactic fragmentation as a stylistic choice of the narrator and preserves the majority of suspension points of the source text. Cases of omission are rare, but there are occasions in which the subtitler omits the fragmentation and completes the interrupted sentence.

The following table depicts some selected examples of preservation of speech fragmentation markers:

SOURCE TEXT (EN)	TARGET TEXT (IT)
“He was talking about democr... Oh, he was just really kind and supportive with my work.” (S01E01, 4:45)	“Stava parlando di democ... Oh, era proprio... troppo gentile e mi incoraggiava nel lavoro.”
“OK, that wasn't my intention, but ... As you are probably aware, we haven't had the opportunity to support many, any women led businesses since the... Sexual harassment case.” (S01E01, 7:39)	“Non era mia intenzione, ma... Come lei probabilmente sa, non abbiamo avuto la possibilità di finanziare molte... anzi, nessuna attività gestita da donne dal... Dalla denuncia per molestie sessuali.”
“So, uh, it-it means a-a great deal to-to both of us that, uh, you, that-that we, the family are-are all together here for, uh ... for a very special family, uh... gang bang, um... Just,	“Allora, un attimo... Per noi due significa molto che voi, noi, la famiglia sia tutta riunita qui per... Per una gangbang di famiglia molto speciale. Anche solo,

you know, b-being here, uh, you know, I-I know we've had our, you know... (S02E01, 3:37)	insomma, il fatto di essere qui... Certo, so che abbiamo avuto i nostri...
--	---

Selected examples of *preservation* of speech fragmentation markers in *Fleabag*.

What can immediately be detected from these examples is that not only do subtitlers preserve accurately the suspension points as in the source text, but they even use them more frequently, which is coherent with the increase resulting from the frequency list analysis of the Italian corpus of the series. The last example in the table above, in particular, deserves special attention as it evidently demonstrates the fragmentation style of the original text is largely maintained, even though specific hesitation markers such as *uh*, *um* and repetitions are systematically eliminated.

The examples in the table below show instances of omission of the suspension points in which the sentence is completed by the translator, or substituted to make the sense clearer.

SOURCE TEXT (EN)	TARGET TEXT (IT)
“Keith, can...” (S01E04, 10:14)	“Keith, per favore.”
“I've told you about ... - Oh my god, I'm totally... I know, I know the Finland promotion. That's amazing.” (S01E04, 16:14)	“Te ne ho parlato! - Oddio... lo so, la promozione, fantastico!”
“Just don't ... talk too much or try to pretend you know anything about the company.” (S02E03, 00:28)	“Non parlare troppo, o prova a fingere di non sapere nulla dell'azienda.”

Selected examples of *omission* of speech fragmentation markers in *Fleabag*.

The motivations behind these omissions are varied. In the first instance, the source utterance implies a request left unfinished: the speaker is trying to ask Keith, who is shouting, to calm down. The request is so obvious that it does not need to be finished. Its rendition with “per favore” replicates the same effect.

In the second example, the overlapping voices and the fast pace of the scene result in a partial omission in order to preserve some form of fragmentation respecting the boundaries

imposed by the spatial constraint of the subtitle format. The last case omits the emphatic pause of the original used to strengthen the warning function of the utterance. In the Italian rendition, the sentence is reproduced conventionally.

Overall, fragmentation is generally preserved by the translator, mostly through suspension points, even though cases of omission are not uncommon. Despite recognising speech fragmentation as a stylistic feature of the source text and part of the characterisation of the protagonist, the translator has to come to terms with the condensation and partial neutralisation caused by the subtitle format, consistent with the standardisation and simplification tendencies in AVT attested in Section 3.2.

3.3 Condensation and Neutralisation: a Discourse Markers analysis

In order to investigate where condensation occurs at the lexical level, attention was turned to discourse markers. As Chaume (2004a) observes, pragmatic elements such as discourse markers are among the most frequently omitted features in subtitle translation, as, in film dialogues, they often carry little semantic content and are, therefore, translators' first candidates for omission due to spatial constraints. A second potential cause of token reduction might be syntactic condensation, that is to say the simplification of complex syntactic structures such as subordinate clauses or verbal phrases, documented by Díaz-Cintas and Remael (2007) as a systematic strategy in subtitle translation. However, even though it should be acknowledged for its contribution to token reduction, a full quantitative analysis of this tendency would require a manual syntactic annotation of the corpus that goes beyond the scope of the present research.

For the purposes of this analysis, the focus will shift to three selected discourse markers: *well*, *I mean*, *you know*. These three items have been selected as they are among the most frequent discourse markers in informal spoken English and appear systematically in the dialogue of both series, functioning as markers of hesitation, involvement and register. For each discourse marker, a two-step concordance search was conducted on both corpora: a basic search to retrieve the total number of instances of each item present in each corpus, followed by a parallel concordance query. A sample of 30 randomly picked instances was manually annotated in order to identify the translation strategy adopted and calculate the omission rate.

The first analysed discourse marker is *well*. A basic concordance research shows that it occurs a total of 1354 times in *The Marvelous Mrs. Maisel*'s corpus and 187 times in *Fleabag*'s corpus. As shown in the following chart, the parallel concordance analysis reveals a systematic tendency towards omission in both series, albeit at different rates.

TRANSLATION STRATEGY	<i>The Marvelous Mrs. Maisel</i>	%	<i>Fleabag</i>	%
Omission	15	50%	11	37%
<i>Beh</i>	10	33%	12	40%
<i>Allora</i>	3	10%	2	7%
<i>Bene</i>	1	3%	1	3%
<i>Sì</i>	1	3%	1	3%
<i>Dài</i>	-	-	1	3%
<i>Niente</i>	-	-	1	3%

Translation strategies for *well* in *The Marvelous Mrs. Maisel* and *Fleabag* (sample: 30 instances per series).

In *The Marvelous Mrs. Maisel*, *well* is omitted in 50% of cases, suggesting that the translator systematically sacrifices this element due to the spatial constraints of the subtitle format. In *Fleabag*, the omission rate is lower at 37%, with *beh* emerging as the preferred equivalent in 40% of the cases. The following examples illustrate the main strategies identified.

SOURCE TEXT (EN)	TARGET TEXT (IT)	SERIES	STRATEGY
“ <i>Well</i> , I hate Moishe and Shirley”. (S03E05, 33:40)	“Odio Moishe e Shirley.”	<i>The Marvelous Mrs. Maisel</i>	Omission
“ <i>Well</i> , I got to see what he does.” (S04E06, 10:24)	“Devo vederlo all’opera.”	<i>The Marvelous Mrs. Maisel</i>	Omission
“ <i>Well</i> , why the fuck did you do that?” (S01E01, 16:02)	“E perché cazzo lo hai fatto?”	<i>Fleabag</i>	Omission
“ <i>Well</i> , I just thought I’d pop by.” (S01E02, 01:18)	“Ho pensato semplicemente di passare.”	<i>Fleabag</i>	Omission

“ <i>Well</i> , at least he’s not living with that girl.” (S01E03, 10:51)	“ <i>Beh</i> , almeno non si è trasferito da quella ragazza.”	<i>The Marvelous Mrs. Maisel</i>	Reformulation
“ <i>Well</i> , I think you’re being really sexy.” (S01E02, 18:19)	“ <i>Beh</i> , penso che sei davvero sexy.”	<i>Fleabag</i>	Reformulation
“ <i>Well</i> , see if you can manage to keep her out of here.” (S01E08, 32:07)	“ <i>Allora</i> seguila lontano da qui.”	<i>The Marvelous Mrs. Maisel</i>	Reformulation
“OK, <i>well</i> , I’ll get a photograph now.” (S01E02, 02:48)	“Ok, <i>allora</i> ... la vado a fotografare subito.”	<i>Fleabag</i>	Reformulation

Selected examples extracted from the parallel concordance of the two study corpora.

The examples presented in this chart illustrate two main strategies: omission and reformulation, used in similar proportions in both corpora. In both the examples presented, *beh* is the most recurrent rendering as it performs a similar pragmatic function in spoken Italian as *well* does in informal spoken English. Reformulation, therefore, replaces one item with another, achieving what Baker (1992) defines as functional equivalence. However, for the sake of this study, omission is the most relevant strategy as it demonstrates the tendency towards reduction and neutralisation central to this analysis.

The difference in rate between the two series, with a much higher reduction in *The Marvelous Mrs. Maisel*, demonstrates a stronger pressure for reduction in a significantly more lexically dense TV show.

Notably, in one instance from *Fleabag*, *well* is omitted when co-occurring with a profanity (“*Well*, why the fuck did you do that?” is rendered as “E perché cazzo l’hai fatto?”, *Fleabag*, S01E01), suggesting a hierarchy in condensation choices whereby the translator prefers to sacrifice the discourse marker in order to preserve the expressive power of the taboo language. This tendency will be further examined in the analysis of profanities in the following section.

The same analysis will now be carried out with the discourse marker *you know*.

TRANSLATION STRATEGY	<i>The Marvelous Mrs. Maisel</i>	%	<i>Fleabag</i>	%
Omission	21	70%	10	33%
<i>Sai</i>	6	20%	9	30%
<i>Sapete</i>	1	3%	-	-
<i>Avete presente</i>	1	3%	-	-
<i>Insomma</i>	1	3%	1	3%
<i>Capisci</i>	-	-	1	3%
<i>Cioé</i>	-	-	1	3%
<i>In effetti</i>	-	-	1	3%
<i>No?</i>	-	-	1	3%
<i>Certo</i>	-	-	1	3%
<i>Lo sai</i>	-	-	1	3%
<i>Vedi</i>	-	-	1	3%

Translation strategies for *you know* in *The Marvelous Mrs. Maisel* and *Fleabag* (sample: 30 instances per series).

This chart demonstrates analogous tendencies to the translation of *well*: omission and reformulation are the main strategies used. However, in this case, the rate of use is very different showing a significantly higher percentage of omission in *The Marvelous Mrs. Maisel* in comparison to *Fleabag*. The chart also highlights *sai* as the main equivalent, even though *Fleabag* offers a wider variety of equivalents for this item as shown in the following examples.

SOURCE TEXT (EN)	TARGET TEXT (IT)	SERIES	STRATEGY
“ <i>You know</i> , the show was not very good.” (S04E03, 51:37)	“Lo spettacolo non era granché.”	<i>The Marvelous Mrs. Maisel</i>	Omission
“He kind of owes us, <i>you know?</i> ” (S04E05, 08:27)	“Ce lo deve.”	<i>The Marvelous Mrs. Maisel</i>	Omission

“ <i>You know</i> , my emotions are up and down.” (S02E02, 20:10)	“Ho le emozioni scombussolate.”	<i>Fleabag</i>	Omission
“Really successful couples do it, <i>you know</i> ?” (S01E02, 17:57)	“Le coppie che funzionano bene lo fanno.”	<i>Fleabag</i>	Omission
“ <i>You know</i> , with that big Yalta head?” (S01E08, 43:05)	“ <i>Sapete</i> , con quel testone alla Yalta?”	<i>The Marvelous Mrs. Maisel</i>	<i>Reformulation</i>
“ <i>You know</i> , the most fascinating thing about Father here...” (S02E01, 05:46)	“ <i>Sapete</i> , la cosa più interessante del nostro Padre...”	<i>Fleabag</i>	<i>Reformulation</i>
“And, <i>you know</i> , I heard that scraping noise...” (S01E07, 00:45)	“E, <i>insomma</i> , avevo sentito qualche scricchiolio...”	<i>The Marvelous Mrs. Maisel</i>	<i>Reformulation</i>
“... sweet little something just to keep it, <i>you know</i> ...” (S01E02, 18:36)	“qualcosina di tenero, giusto per alimentare... <i>capisci</i> ?”	<i>Fleabag</i>	<i>Reformulation</i>

Translation strategies for *you know* in *The Marvelous Mrs. Maisel* and *Fleabag* (sample: 30 instances per series).

Observing these samples, the data suggests a tendency towards omission when *you know* is positioned at the end of the sentence, whereas when positioned in the middle of an utterance it tends to be preserved and rendered with a functional equivalent. In initial position, the pattern is less consistent. However, in some instances, the omission of discourse markers co-occurs with the omission of other elements, as is the case for the second example, in which *kind of* is omitted as well. Therefore, while the data points to a general tendency, a more systematic qualitative analysis at utterance level would be necessary in order to confirm such a claim.

Finally, in order to demonstrate systematicity, one last discourse marker will be analysed: *I mean*.

TRANSLATION STRATEGY	<i>The Marvelous Mrs. Maisel</i>	%	<i>Fleabag</i>	%
Omission	18	60%	11	37%
<i>Cioè</i>	4	13%	4	13%
<i>Voglio dire</i>	3	10%	1	3%
<i>Insomma</i>	3	10%	3	10%
<i>Sul serio</i>	-	-	2	7%
<i>Anzi</i>	1	3%	-	-
<i>Che</i>	1	3%	-	-

Translation strategies for *I mean* in *The Marvelous Mrs. Maisel* and *Fleabag* (sample: 30 instances per series).

As shown in this overview chart, the pattern is confirmed: omission strategies are more frequent in *The Marvelous Mrs. Maisel* (60%) in comparison to *Fleabag* (37%), consistently with what was shown in the other charts. Interestingly, both series prioritise *cioè* as the main reformulation, whereas the most literal translation *voglio dire* only appears in the samples taken from the first series. This could suggest a more literal translation strategy, a pattern to look for in further analysis.

The present table illustrates the main translation strategies identified.

SOURCE TEXT (EN)	TARGET TEXT (IT)	SERIES	STRATEGY
" <i>I mean</i> , the flu is very serious." (S01E02, 26:35)	"Sono cose serie."	<i>The Marvelous Mrs. Maisel</i>	Omission
" <i>I mean</i> , two months away is gonna kill our momentum." (S02E04, 05:29)	"Due mesi via ci faranno perdere slancio."	<i>The Marvelous Mrs. Maisel</i>	Omission

“ <i>I mean</i> , Elaine is being amazing.” (S02E02, 20:16)	“Elaine è stata straordinaria.”	<i>Fleabag</i>	Omission
“ <i>I mean</i> , what do you think?” (S01E02, 18:17)	“Cosa ne pensi?”	<i>Fleabag</i>	Omission
“ <i>I mean</i> , how are your children even alive...” (S01E03, 09:22)	“Cioè, come fanno ad essere vivi i tuoi figli...”	<i>The Marvelous Mrs. Maisel</i>	<i>Reformulation</i>
“ <i>I mean</i> , there’s nothing wrong...” (S01E04, 14:56)	“Cioè, non c’è niente...”	<i>Fleabag</i>	<i>Reformulation</i>
“ <i>I mean</i> , they're your friends as much as Joel's.” (S01E03, 11:24)	“Voglio dire, sono sia tuoi amici che di Joel.”	<i>The Marvelous Mrs. Maisel</i>	<i>Reformulation</i>
““ <i>I mean</i> , where the fuck even are they?”” (S01E02, 13:52)	“Insomma, ma che cazzo dovrebbero essere, poi?”	<i>Fleabag</i>	<i>Reformulation</i>

Translation strategies for *I mean* in *The Marvelous Mrs. Maisel* and *Fleabag* (sample: 30 instances per series).

Through the analysis of these discourse markers, a systematic pattern of omission is shown in both TV series, with a significantly higher rate in *The Marvelous Mrs. Maisel* in comparison to *Fleabag*. As illustrated at the beginning of this section, discourse markers are indeed part of the cause of the notable token reduction measured through frequency lists. This tendency is consistent with the simplification and condensation patterns documented in AVT research by Pavesi (2019) and Baker’s (1993) universals of translation: when discourse markers carry little propositional content, they tend to be sacrificed under spatial constraints, whereas they are preserved and translated when deemed relevant in the narrative context.

Looking at it through Arias-Badia’s (2015) lens, this systematic omission tendency can be interpreted as a form of character neutralisation: indeed, by eliminating discourse markers

characterising the protagonists' speech, the Italian subtitles inevitably produce a more neutral target text that partially erases the distinctiveness of each voice.

3.4 *Fleabag*: the Case of Profanities

The present section turns to the third identified cause: the possible omission of marked vocabulary, that is to say slang and profanities. More specifically, in this study a focus will be put on profanities in *Fleabag*. Before diving deep into the parallel concordance analysis, a keyword analysis is considered necessary in order for these terms to emerge as systematically relevant to the study of *Fleabag*'s corpus. Through the Keyword tool in Sketch Engine, the DIY corpora can be compared to a reference corpus so as to identify keywords, terms and n-grams typical of the selected corpus. For the purposes of this study, the reference corpus used is the British National Corpus (BNC), more specifically, the Spoken subcorpus will be employed, as subtitling aims at simulating spoken language, rather than written. Furthermore, not only does spoken BNC include informal conversations, but it is also a British corpus, coherent with *Fleabag* being a British TV series. As addressed in Chapter 2, the results are ranked according to the Log-Likelihood score, with the terms more statistically relevant to the comparison to the reference corpus appearing at the top.

Among the categories that emerged from the keyword analysis, particular attention will be paid to the category of profanities and taboo language, as these constitute the primary dimension of marked language in *Fleabag* and are directly relevant to the investigation of neutralisation in the Italian rendering of the subtitles. Having identified the main taboo words in the corpus, a parallel concordance query is carried out. It should be noted that the present analysis examines the Italian renditions of English swearwords identified through lemma searches in the source corpus, without accounting for possible cases of compensation in the target text.

The first word to analyse is *fuck* with its different forms (*fuck, fucking, fucked, fucker*), therefore, using the query type “*lemma*”. Out of the 171 total instances generated, a sample of 50 randomly selected instances shows an 84% tendency towards direct equivalents of the word, while drawing on a wider variety of options in Italian (*cazzo, fanculo, vaffanculo, fottuto, scopare, di merda*). This result appears as contradictory to the neutralisation hypothesis: in this instance, it is evident that neutralisation does not appear to be a universal tendency, but rather the subtitler chooses to preserve the force of the original word. This tendency may reflect the centrality of profane language to the series' characterisation, for which its omission or attenuation would amount to neutralising the impact of the series.

The following selected examples illustrate the range of translation strategies identified, from the most commonly found direct equivalences to the rare cases of reformulation, neutralisation or omission.

SOURCE TEXT (EN)	TARGET TEXT (IT)	STRATEGY
“ <i>Fucking</i> last minute bastard trendy party.” (S01E01, 13:32)	“ <i>Fottutissime</i> feste alla moda di merda dell’ultimo minuto.”	Direct equivalent
“And I <i>fucked</i> my friend by <i>fucking</i> her boyfriend.” (S01E06, 22:44)	“E ho <i>distrutto</i> la mia amica, <i>scopandomi</i> il suo ragazzo.”	Neutralisation + Direct equivalent
“He was really <i>fucking</i> affectionate.” (S01E01, 06:15)	“In più era <i>davvero</i> molto affettuoso.”	Neutralisation
“Wait, what the <i>fuck</i> is going on here?!” (S02E06, 07:33)	“Aspettate, cosa sta succedendo qui?”	Omission

Selected examples of translation strategies for *fuck* in *Fleabag*.

In the first case, the subtitler not only chooses a direct equivalent translating *fucking* with *fottutissime*, but also adds a superlative form as an intensifier. In the second example, the first *fucked* is rendered with the verb *distruggere*, which captures the figurative meaning of causing harm or betraying someone in the original utterance, rather than its marked register. While the propositional meaning is prioritised and preserved, not only is there a neutralisation of the original profane register, but there is also a loss in the repetition of the word *fuck* in its two occurrences. A more faithful rendering that would preserve both the profanity and the phonetic repetition could have been “E ho fottuto la mia amica, fottendomi il suo ragazzo”. The translation *scopandomi*, however, is still a perfect direct equivalent.

The third case presents an example of neutralisation with the profanity rendered as a standardised *davvero* replacing the taboo language. It should be noted, however, that in this instance, *fucking* functions as an intensifier, a use that could be rendered in Italian with *fottutamente*. Nevertheless, this equivalent tends to sound less natural in Italian than in English. The fourth case presents a total omission of the word.

The next word to analyse following the same procedure is *shit*. Using the parallel concordance tool with the lemma query *shit*, 23 total occurrences appear. As it is a small

number, all 23 occurrences will be observed. As a result of this observation, it appears once again that the dominant strategy used by the subtitler is equivalence (around 78%). In other instances, neutralisation is used, namely when *shit* appears within an idiomatic expression. In the following chart, these examples will be observed and commented.

SOURCE TEXT (EN)	TARGET TEXT (IT)	STRATEGY
“Oh, <i>shit</i> !” (S01E02, 10:48)	“Oh, <i>merda</i> !”	Direct Equivalent
“Oh <i>shit</i> !” (S01E02, 21:20)	“ <i>Cazzo</i> !”	Functional Equivalence
“ <i>Shit</i> got real?” (S01E02, 12:53)	“Le cose si fecero serie?”	Neutralisation
“Either she's got her period or some serious <i>shit</i> 's gone down.” (S01E05, 07:07)	“O le è venuto il ciclo oppure sta succedendo <i>qualcosa</i> di tremendo.”	Neutralisation
“I’m just gonna go for a <i>shit</i> .” (S02E05, 00:39)	“Vado a <i>cagare</i> .”	Direct Equivalent

Selected examples of translation strategies for *shit* in *Fleabag*.

The first two examples show two different occurrences of *shit* with different renditions in Italian: the first one is a direct equivalence with the literal translation of *shit* as *merda*; the second one can be considered a functional equivalence as the translator chooses a different profanity in Italian (*cazzo*), still preserving the same pragmatic effect.

In the third example, as anticipated before, the profanity is lost in translation as it is part of the slang expression *shit got real*. This phrase is normally used when a situation escalates dramatically and becomes a serious matter. The translator uses a neutral expression in the target language (*le cose si fecero serie*) to transfer the same meaning, but loses the force of the profanity. Upon consulting the official subtitle translation made for Amazon Prime Video, “è diventata seria?” was chosen to render the sentence. In the context of the scene¹⁸, a more faithful rendering preserving the profanity could have been “la merda si è fatta seria?” or “si mise di

¹⁸ Fleabag S01E02, 12:25.

merda?” or even “è diventata una merda?”. Another example of neutralisation happens in the following example when the subtitler renders “*some serious shit’s gone down*” as “*sta succedendo qualcosa di tremendo*”¹⁹. The official translator uses a similar rendition choosing: “*O ha il ciclo, o sta succedendo qualcosa di molto importante.*” A more faithful translation, could have been “*o ha il ciclo, o è successo un casino.*” In this case, a slang expression of spoken Italian is chosen (*casino*), but the profanity is still omitted. In order to keep it an expression such as “*è successo un casino del cazzo*” could also be used, less literal as a functional equivalent (*cazzo*) is used instead of the literal translation *merda*.

Finally, the last example presented can be considered a direct equivalent, even though in Italian it sounds more vulgar and colloquial in comparison to its English original. Further terms identified through the keyword analysis, such as *shit* and *asshole*²⁰, confirm the same pattern of direct equivalence as the dominant strategy, with a tendency towards neutralisation limited to specific idiomatic contexts.

Overall, the analysis of profanity in *Fleabag* reveals a pattern that partially contradicts the neutralisation hypothesis proposed by Arias-Badia (2015) and of standardisation and simplification documented in AVT research (Pavesi, 2019): rather than systematically softening the taboo language, the translator tries to consistently preserve its pragmatic force through direct or functional equivalents. The exceptions identified suggest that neutralising profanity is not a systematic choice, but rather a response to specific constraints. This inclination could be interpreted in the light of the centrality of profanity to the series’ characterisation: strong language in *Fleabag* constitutes a defining feature of the protagonist’s voice and identity, often reprimanded for being inappropriate. In this instance, the subtitler’s choices are driven by the specific linguistic identity of the series, rather than by general subtitling conventions.

3.5 *Fleabag*: the Translation of Marked Language

Following Díaz-Cintas & Remael’s (2007) classification of translation strategies for marked speech discussed in Chapter 1, this section is dedicated to the rendition of slang and its non-standard grammar and specific lexical features. Slang is not necessarily taboo language, but it is indeed culturally marked: in this case, it is namely middle-class British slang. However, contrary to swearwords, slang cannot be considered a closed category and it does not lend itself

¹⁹*Fleabag* S01E05, 6:55.

²⁰ *Shit* occurs 15 times, always translated as *troia* or *puttana*, whereas *asshole* only has 3 occurrences, all rendered as *stronzo/a*. Only in one case, it is rendered as *stronzetta*, adding a nuance of irony absent in the original.

to automatic corpus querying. For this reason, the approach used will be a corpus-assisted type. Slang expressions have been manually detected and extracted from the TV show and corpora will be used in order to easily identify how they have been rendered. Furthermore, a direct comparison between the fansubs and the official Amazon Prime Video subtitles will be carried out for selected significant cases. The idiomatic British expressions will be analysed simply by querying the exact string into the parallel concordance research tool in order to observe how it was translated. Furthermore, Sketch Engine matches each line with the document it was taken from, making it easier to retrieve the episode the line was in and look for the official translation on the streaming platform.

Slang and idiomatic expressions pose a cultural equivalence challenge. For this reason, the classification of translation strategies adopted here follows Díaz-Cintas and Remael's (2007) taxonomy for marked speech discussed in Chapter 1 (calque, loan, explication, substitution, transposition, lexical recreation, compensation, omission, addition).

For easier research, the slang identified could be divided into three categories:

1. idiomatic British expression: *can't be arsed; she blows hot and cold; pop to the loo; I'd put a tenner on it; to jump ship; easy, tiger; gives me some edge; a cheap thrill; time to throw the net out; we are a little bit tired on the purse strings; we're just going to do a quick whizz around; I'm such a tool box; tat; cracking.*
2. British cultural realia: *I had to do a flash poo in Pret; G&T from M&S; Tooting;*
3. Wordplays and jokes: *Do you know what the lesbian app for Grindr is called? Twat Now; Sexhibition; weaky/strongy; let's go fucking crazy vs. let's go fuck like crazy; swallow a short jail term; too much stuff, absolutely no staff.*

3.5.1 Fleabag: Rendering Idiomatic British Expressions

ST (EN)	FANSUB (IT)	OFFICIAL (IT)	STRATEGY
"I <i>can't be arsed</i> , darling." (S02E03, 14:39)	" <i>Non mi va</i> , tesoro."	" <i>Non mi va</i> , tesoro."	Substitution
"Well, I have a guinea pig, but <i>she blows hot and cold.</i> "	"Ho un porcellino d'India, <i>ma è una tipa scostante.</i> "	"Ho un porcellino d'India, <i>ma è una tipa scostante.</i> "	Substitution

(S02E02, 15:52)			
“Honestly, you could tell him you’re going to <i>pop to the loo</i>, ...” (S01E03, 6:04)	“Sul serio, se gli dicessi che <i>vai un attimo in bagno</i>, ...”	“Magari gli dici che devi <i>andare in bagno</i>, ...”	Explicitation
“I’d put a <i>tenner</i> on it.” (S01E02, 1:34)	“Scommetto <i>dieci sterline</i>.”	“Ci scommetto <i>10 sterline</i>.”	Explicitation
“It’s the kid’s choice if it wants to <i>jump ship</i>, right?” (S02E01, 22:38)	“É una scelta del piccolo se vuole <i>abbandonare la nave</i>, giusto?”	“É una scelta del piccolo se vuole <i>abbandonare la nave</i>, giusto?”	Calque
“Whoa, <i>easy</i>, <i>tiger</i>.” (S01E03, 10:55)	“<i>Vacci piano</i>, <i>tigre</i>.”	“<i>Vacci piano</i>, <i>tigre</i>.”	Calque
“<i>Gives me some edge</i>.” (S02E02, 2:34)	“<i>Mi dà l’aria da duro</i>.”	“<i>Mi dà l’aria da duro</i>.”	Substitution
“Oh, just a really, really <i>cheap thrill</i>.” (S01E03, 13:53)	“Oh, giusto un <i>brivido molto poco costoso</i>.”	“Un <i>brivido a buon mercato</i>.”	Calque / Transposition
“Time to <i>throw the net out</i>.” (S01E03, 5:10)	“Adesso vediamo <i>chi abbocca</i>.”	“É ora di <i>gettare l’esca</i>.”	Transposition / Calque
“We are just a little bit <i>tight on the purse strings</i> too.” (S01E05, 10:23)	“Capirai che anche noi <i>stiamo tirando la cinghia</i>.”	“Anche noi <i>siamo in ristrettezze</i>.”	Transposition / Substitution

“We’re just going to do a <i>quick whizz around</i> to show our faces.” (S01E06, 11:51)	“Facciamo solo <i>un giro veloce</i> - per far vedere che c’eravamo.”	“Diamo solo <i>una rapida occhiata</i> per farci vedere.”	Substitution
“I mean, don’t get me wrong, <i>I’m chuffed to my boots.</i> ” (S01E03, 12:52)	“Non fraintendermi, <i>faccio i salti di gioia.</i> ”	“Non fraintendermi, <i>sono piacevolmente sorpreso.</i> ”	Transposition / Substitution
“I’m such a <i>tool box.</i> ” (S01E03, 13:02)	“Sono proprio un <i>disastro.</i> ”	“Sono un vero <i>idiota.</i> ”	Substitution
“Sorry about all the <i>tat.</i> ” (S02E02, 4:06)	“Scusa per la <i>confusione.</i> ”	“Scusa per la <i>confusione.</i> ”	Substitution
“I don’t want to boast, but I make a <i>cracking...</i> ” (S02E02, 3:44)	“Non per vantarmi, ma <i>so fare miracoli col...</i> ”	“Non per vantarmi, ma <i>so fare miracoli col...</i> ”	Substitution

Idiomatic British expressions in *Fleabag*: fansub vs. official translation strategies.

Observing the present chart, it is evident that the dominant strategy is substitution with a standard term of easier comprehension. In most of these cases, the translation is consistent in both fansubs and official subtitles showing a standardised translation strategy. Nevertheless, the analysis of slang expressions becomes relevant in some instances that will be now discussed in detail. A notable case is the expression “*Oh, just a really, really cheap thrill*” in which a *cheap thrill* indicates “a short-lived moment of excitement or pleasure, esp. one that is ultimately unsatisfying” (Oxford University Press. (n.d.). Cheap thrill, n. In Oxford English dictionary). In the fansub version it is rendered with a literal translation word for word (*cheap* = *poco costoso*; *thrill* = *brivido*). In the official version, instead, the translator looks for an Italian idiomatic expression that transposes the same meaning and uses the expression *a buon mercato*. Both solutions successfully convey the original meaning.

For the scope of this analysis, a further query was conducted on a reference subtitles parallel corpus in order to verify the dominant translation adopted for the expression in subtitling in general. The reference corpus used here is the *OpenSubtitles 2018 parallel – English* corpus. Querying the lemma *thrill* into the parallel concordance search tool and filtering it with *cheap* on the left side of the word, 32 co-occurrences appear. The expression *a buon mercato* is one of the most recurrent, yet some other options that could have been used in this instance are: *un brivido da due lire*, *divertimento da quattro soldi*, *emozione da quattro soldi*, *brividi facili*, *emozione a basso costo*, *emozioni passeggiare*, *emozioni da poco*.

The expression *time to throw the net out* derives from the fishing field. In fact, by querying it into the BNC reference corpus and looking for concordances of the lemmas *throw* and *net*, they always occur together when talking about catching a fish. Therefore, in both cases fansubber and official subtitler have maintained the fishing reference, even though the official version can be considered a more literal calque as it describes the act of throwing something in the water in order for the prey to take the bait. In the OpenSubtitles corpus as well, it appears mostly with a literal meaning, yet there are some instances of figurative meaning as well: namely, it is also translated as *tendere una trappola*, *stanare* or *sparare nel mucchio*. Of these options, the first one would also be a good rendition of the idiomatic expression in this context.

Different strategies are adopted for the idiomatic expression *tighten the purse strings* meaning “to make funds or the supply of money less (or more) readily available” (Webster’s New World College Dictionary, 5th Digital Edition. Copyright © 2025 HarperCollins Publishers). The fansubber relies on a transposition with a different idiomatic expression in Italian (*tirare la cinghia*), reproducing the same meaning²¹ and imagery; whereas the official subtitler substitutes the expression with a more formal *in ristrettezze*. This could also be defined as a case of neutralisation (Arias-Badia, 2015) in which an idiomatic colourful expression is rendered in a more neutral form. In this instance, the fansubber’s choice appears to be more coherent with the register of the series. To verify this hypothesis, the reference corpus OpenSubtitles was used once again (parallel concordance query of *tight*, filtered with *string* as co-occurrence), but no occurrences of the phrase as an idiomatic expression appear. Consequently, the same query was typed in the BNC reference corpus where the figurative expression appears mostly in phrases related to the world of affairs. Therefore, while *in ristrettezze* could still be considered an accurate rendition with respect to the field of use of the original expression, it suggests a degree of neutralisation of the colloquial tone of the series.

²¹ “to endure hunger or other hardships caused by food shortages or economic restrictions” (author’s translation from Treccani online dictionary, Istituto della Enciclopedia Italiana, n. d.).

One last expression to take under analysis is *to be chuffed to one's boots*. The slang adjective *chuffed* on his own means “pleased, satisfied” (Oxford English Dictionary, “chuffed (adj.),” December 2024) and it is intensified by the addition of *to one's boots*, which is a variation of *chuffed to bits*, UK slang for pleased. On the one hand, the fansub version proposes a transposition with an equivalent idiomatic expression in Italian (*fare i salti di gioia*) that perfectly preserves the colloquial register and the expressive force of the original. The official version, on the other hand, can be considered a substitution or a neutralisation as with *piacevolmente sorpreso* the register changes towards a less colloquial form. Querying the expression *chuffed* on the OpenSubtitles corpus different options are given; among these another accurate alternative for transposition might have been *felice come una Pasqua*.

3.5.2 *Fleabag*: Rendering British Realia

After analysing the idiomatic expressions, attention is now turned to the second category identified: that of British realia as culture-bound terms. In the case of *Fleabag*, the realia extracted from the text are mostly geographically bound.

BRITISH REALIA	FANSUB (IT)	OFFICIAL (IT)	STRATEGY
“I had to do a flash poo in <i>Pret</i> .” (S01E01, 9:18)	“Ho dovuto fare la cacca al <i>ristorante</i> .”	“Ho fatto la cacca al volo da <i>Pret</i> .”	Explicitation / loan
“I’ve got cans of <i>G&T</i> from <i>M&S</i> .” (S02E02, 4:23)	“Ho delle lattine di <i>gin tonic</i> .”	“Ho delle lattine di <i>gin tonic</i> .”	Substitution
“ <i>Tooting</i> .” (S02E02, 7:50)	“Assurdo.”	“Assurdo.”	Omission

British realia in *Fleabag*: fansub vs. official translation strategies.

These three cases are extremely interesting, each for different analytical reasons. Looking at the first instance, the fansubber uses an explicitation (*ristorante*) generalising and losing the specific reference to the brand *Pret* or *Pret à Manger*, a British café chain. Despite being less widespread, a few shops are present in Italy making the brand recognisable to at least part of the Italian audience. Therefore, the official version using a loan to maintain the realia, can be considered a more faithful translation strategy. In support of this observation, when querying

the word *Pret* on the OpenSubtitles reference corpus, the brand name is preserved in the majority of the Italian renditions.

A different case is the rendition of the double realia G&T and M&S. G&T is an abbreviation for the drink *Gin and Tonic*, whereas M&S stands for *Marks & Spencer*, a well-known British department store. This reference is culturally bound to a very specific milieu, yet both versions coincide in using a substitution strategy that omits the abbreviations and the realia. Opting for a loan here, as in the previous example, would have preserved the cultural specificity, but the communicative effect would have been different as M&S are not well-known in the Italian culture and G&Ts are commonly called *gin tonic*. Nevertheless, a degree of cultural loss is inevitable, as the specific social connotations of the source text, evoking the British middle-class milieu, are not recoverable through this substitution.

Finally, *Tooting* is a geographical realia, which is completely misinterpreted in both renditions. Both the fan-made and the official renditions adopt *assurdo* as the translation, which appears completely detached from the source text. This is one of those cases in which the rendition can only be faithful, if analysed in its narrative and audiovisual context. The word is pronounced by Claire as she is sitting outside her sister's café during Chatty Wednesday (*"If you buy something, you have to have a chat with someone you don't know"* *Fleabag*, S02E02 7:20). A man walks up to her, asking her where she is from, but she is unsettled by the situation and mumbles that she does not intend to answer; however, as the man does not walk away, she blurts out *"Tooting"*, apparently out of context. Actually, Tooting is the name of a district in South London. Both translators have missed the reference, resulting in the involuntary omission of the realia and loss not only of the geographical reference, but also of the comic effect of the original. Two options worth considering so as to produce a more faithful translation might have been preserving the original realia, or opting for a generalisation such as *Londra sud*, which would have maintained the geographical dimension and the irony, as well as reduced the opacity of the source reference.

3.5.3 *Fleabag*: Rendering Wordplays and Jokes

The third pinpointed group of marked language expressions concerns wordplays. Referencing Hanks's CPA (2012) extensively discussed in Chapter 2, wordplays can be classified as exploitations, that is to say deviations from regular patterns of use so as to produce a creative or humorous effect, in contrast to norms, systematically detectable in a large corpus.

SOURCE (EN)	FANSUB (IT)	OFFICIAL (IT)	STRATEGY
“Do you know what the lesbian app for Grindr is called? <i>Twat-Nav.</i> ” (S01E04, 00:46)	“Lo sai come si chiama la versione per lesbiche di Grindr? <i>Naficatore.</i> ”	“Sai cosa usano le lesbiche per incontrarsi? <i>Fica-advisor.</i> ”	Lexical recreation
“It’s a <i>sexhibition.</i> ” (S01E05, 10:31)	“É una <i>sesposizione.</i> ”	“É una <i>sexibizione.</i> ”	Lexical recreation
“A <i>weaky.</i> ” [...] “You’re a... <i>strongy.</i> ” (S02E02, 22:08)	“Un <i>pappamolle.</i> ” [...] “Tu sei una <i>pappaforte.</i> ”	“Un <i>pappamolle.</i> ” [...] “Tu sei una <i>pappaforte.</i> ”	Lexical recreation
“Let’s go <i>fucking crazy</i> tonight, then.” “What did you say?” “I just said, “Let’s go <i>fuck like crazy</i> tonight.” (S02E03, 5:10)	“Allora, <i>scoliamo</i> di brutto stasera.” “Cosa hai detto?” “Ho detto: “ <i>scopiamo</i> di brutto, stasera.”	“Allora, <i>scoliamo</i> di brutto stasera.” “Cosa hai detto?” “Ho detto: “ <i>scopiamo</i> di brutto, stasera.”	Compensation or lexical recreation
“Well, if you <i>spit</i> guilty, you’ll have to <i>swallow</i> a short jail term.” (S02E02, 11:52)	“Se ti <i>dichiari</i> colpevole, dovrai <i>farti</i> un breve periodo di prigionia.”	“Se ti <i>dichiari</i> colpevole, dovrai <i>farti</i> un breve periodo di prigionia.”	Omission
“So much <i>stuff.</i> Absolutely no <i>staff.</i> ” (S02E02, 4:11)	“Tanta <i>roba</i> da fare, ma niente <i>personale.</i> ”	“Tanta <i>roba</i> da fare, ma niente <i>personale.</i> ”	Omission

Wordplays in *Fleabag*: fansub vs. official translation strategies.

The present chart illustrates a series of analytically interesting creative exploitations. The first pun plays on a parody of the Grindr app for gay men, proposing a name for its lesbian version: Twat-Nav, wordplay from Sat-Nav (satellite navigation), substituting *sat* with *twat* (vulgar term referring to female genitalia), two words that also produce a similar phonetic effect. In their version, fansubbers reproduce exactly the same mechanism: *Naficatore* comes from the fusion

between *navigatore* and *fica* (Italian vulgar term referring to female genitalia) and perfectly recreates the original. The option chosen by the official subtitler loses the reference to the satellite navigator, playing instead on the word TripAdvisor, a travel review platform. The original reference, therefore, is substituted with another perhaps considered more recognisable by the audience. As a consequence, both creative solutions can be considered effective: one privileges faithfulness (foreignising), the other recognisability (domesticating).

Lexical recreation is also at use for the rendition of the neologism *sexhibition*, a fusion between *sex* and *exhibition*, used to describe a sex-themed art exhibition. Contrary to the previous example, in this case it is the fansub version that privileges a more domesticating lexical recreation with the completely Italianised neologism *sesposizione* (*Sesso + esposizione*). The official version, on the contrary, adopts a more foreignising approach as it creates *sexibizione*, which maintains the English prefix *sex* from the source text. Both solutions are effective, though *sesposizione* sounds more natural in Italian.

Another example of lexical recreation can be observed with the opposition *weaky/strongy*, a morphological wordplay creating by adding the suffix *-y* to the adjectives *weak* and *strong*. In this case, both versions adopt a contrast between the already existing Italian word *pappamolle*, which perfectly encapsulates the childish colloquial register of the source text, and its symmetric neologism *pappaforte*, created as a counterinsult using the same mechanism as the original. A further case under analysis is that between the phonetically similar *let's go fucking crazy* vs. *let's go fuck like crazy* perfectly rendered into Italian using the same device: *scoliamo/scopiamo* preserves the phonetic game, as well as the meaning of the original, producing an equivalent humorous effect.

On other occasions, jokes and wordplays can become more difficult to restore. It is, for instance, the case of the wordplay with *spit* and *swallow*. Context is pivotal here: as the lawyer pronounces the sentence using *spit* (meaning “to declare guilty”) and *swallow* (meaning “to accept the sentence”), he is visually spitting and swallowing food voraciously. This double level, visual and verbal, creates the humorous effect of what Zabalbeascoa (1996) defines a complex joke. In Italian, the wordplay is rendered with a neutral form, omitting the verbal game, as well as the connection to the visuals. Nevertheless, this case also illustrates the limitations of the monomodal corpus discussed in chapter 2: the visual dimension of the joke cannot be accessed through subtitles alone, but can only be recovered through direct observation of the audiovisual product.

The last example “*So much stuff, absolutely no staff*” builds on the phonetic similarity between the two semantically different nouns *stuff* and *staff*, almost homophones. In Italian, the

similarity is completely lost and neutralised in both versions. A more faithful solution to at least try to partially reproduce the repetitive sound of the sentence could have been “tanto da fare, nessuno che lo faccia”, even though it does not fully replicate the contrast between the two original terms, demonstrating the untranslatability of purely phonetic wordplay.

3.5.4 *Fleabag*: Concluding Remarks on Marked Language

From the present analysis of marked language in *Fleabag* crucial general tendencies have emerged. Having classified marked language into three main categories, that is to say idiomatic expressions, realia and wordplays, it can be assessed that each group presents its own remarkable patterns of translation, yet the options chosen by the subtitlers are variable.

First and foremost, idiomatic expressions privilege substitution as the dominant translation strategy. However, interesting cases of explicitation, calque and transposition demonstrate that both fansubbed and official versions offer effective solutions, despite the presence of instances of neutralisation, more or less unavoidable, which, in turn, create space for the proposition of alternative renditions. Secondly, for the rendition of realia strategies are more diversified with cases of explicitation, loan, substitution and even omission due to misunderstanding made by the translator. Lastly, when dealing with wordplays and jokes, translators give space to their creativity with their own lexical recreations of the original puns in order to reproduce the same humorous effect as the source text. However, instances of omission due to a lack of equivalents in the target language are present as well.

The contrast between the subtitles translated by fans and those made by official subtitlers sets the foundations for a systematic pattern of translating approach. Fansubs show a predisposition towards more creative solutions that privilege faithfulness to the source text and a more foreignising approach. On the contrary, Amazon’s Prime Video’s official subtitles generally prioritise more neutral and domesticating translation approaches. However, exceptions can be found in both occurrences, as shown in the present analysis.

Overall, the wide variety of translation strategies adopted demonstrates the complexity of translating marked language and all its distinctive features. As professional translators understand the expressive, narrative and lexical force of this heterogeneous category, they do not limit their activity to a systematic pattern, but realise the importance of analysing each instance in its specific context. In this case, looking at the general translation pattern through a corpus-assisted analysis, becomes useful to find the best suited rendition and avoid opaque translation choices.

3.6 *The Marvelous Mrs. Maisel*: Rendering Yiddish Realia

When talking about marked discourse above, a section was dedicated to realia. In the second TV series under analysis here, *The Marvelous Mrs. Maisel*, realia plays an even more pivotal role as the series is set in a very specific milieu with its own cultural, geographical and religious characteristics. The series is set within the Ashkenazi Jewish community of the Upper West Side and the Catskill resorts in which a Yiddish inflected variety of American English (Jewish English) functions as a marker of group identity and cultural specificity with a systematic presence of Yiddish-derived lexical items in the dialogues.

Italian has no direct equivalent with Yiddish (Jewish English) and there is no community specific version of Italian that functions in the same way as a marker of ethnic and cultural belonging for a minority group. The translator is met with different options in order to render these instances. The main duality is between comprehensibility, with domestication strategies that tend to erase cultural specificity and, on the contrary, cultural fidelity, foreignising strategies that opt to preserve the Yiddish at the cost of opacity. Occasionally, intermediate strategies such as glossing, adaption or compensation can be adopted. However, in this specific case, fast-paced dialogue leaves little to no room for glossing or further explanation due to subtitles' specificities and constraints, which can also be culturally responsible for the choice between domesticating and foreignising strategies.

In order to identify the Yiddish terms in the corpora a keyword analysis was conducted using the British National Corpus – Spoken Part as a reference corpus. As explained in Chapter 2, keyword analysis helps identifying the vocabulary that is distinctive of the corpus under study. Therefore, other than toponyms and characters' names, the keyword analysis yielded a list of the main Yiddish terms used in the series. The extracted terms were subsequently organised into three analytical categories based on Florin's (1993) classification mentioned in Section 1.6.

Following this distinction, the first category is that of geographic realia. As briefly mentioned above, the keyword analysis listed a significant number of toponyms²² such as names of neighbourhoods and towns in the specific New York area in which the show is set. These realia, while culturally relevant, are systematically preserved into Italian and are consequently not examined in detail here. The second category under examination will be that of ethnographic realia with terms such as *bagel*, *pastrami*, *kugel*, *latke*, *blintz*, *tush*, *schlep*, *schmoze* (food and lifestyle), *rabbi*, *yom kippur*, *mitzvah*, *mezuzah*, *hannukah*, *torah*, *aliyah*,

²² Find here some instances: *Catskill*, *Hoboken*, *Spokane* (town in New York).

mohel (religion) and *spiel*, *Yiddish* (culture). Finally, socio-political realia constitute the third group of realia with examples such as *schmuck*, *yenta*, *schnorrer*, *shiksa*, *goy*, *chutzpah*, *oy vey*, *mazel tov*, *kvetch*, *kvell*, *bupkis*.

Before diving into the actual analysis, a remark should be made: the transliteration of Yiddish into English generates different orthographic forms of the same terms. As a consequence, when querying each realia into Sketch Engine not all occurrences could appear if the same terms is not presented in the same orthographic forms throughout the corpus. In order to address this issue CQL should be used including the main orthographic variations: for example, `[word="schmuck|shmuck"]`. Using this formula, all occurrences of the word, whether written as *schmuck* or *shmuck* will appear through one single query.

Having identified the specific list of Yiddish realia mentioned in the TV show, a summary table will follow with an overview of the translation strategies adopted. Moreover, a detailed analysis will be carried out for each category, examining the main rendition strategies adopted and investigating interesting instances of creative renditions. In order to classify the translation strategy, the analysis will be referencing Aixelà's (1996) framework explored in Section 1.6.

REALIA	OCCURRENCES	TRANSLATION STRATEGY	RENDITION
<i>Rabbi</i>	43	Linguistic translation	<i>Rabbino</i>
<i>Yom kippur</i>	12	Repetition	<i>Yom kippur</i>
<i>Mitzvah</i>	12	Repetition / absolute universalisation	<i>Mitzvah</i> / <i>precetto</i>
<i>Mezuzah</i>	8	Repetition	<i>Mezuzah</i>
<i>Hannukah</i>	9	Repetition / deletion	<i>Hanukkah</i> / <i>chanukkah</i>
<i>Torah</i>	5	Repetition	<i>Torah</i>
<i>Aliyah</i>	4	Repetition	<i>Aliyah</i>
<i>Mohel</i>	4	Repetition / absolute universalisation	<i>Mohel</i> / <i>circoncisione</i>
<i>Bagel</i>	11	Repetition	<i>Bagel</i>

<i>Pastrami</i>	6	Repetition	<i>Pastrami</i>
<i>Kugel</i>	8	Repetition / limited universalisation / deletion	<i>Kugel / pudding</i>
<i>Latke</i>	1	Deletion	<i>Cocktail</i>
<i>Blintz</i>	5	Repetition	<i>Blintz</i>
<i>Tush</i>	13	Absolute universalisation	<i>Chiappe / sedere / culetto</i>
<i>Schlep</i>	1	Absolute universalisation	<i>Visitare</i>
<i>Schmooze</i>	1	Limited universalisation	<i>Attaccare bottone</i>
<i>Spiel</i>	1	Linguistic translation	<i>Discorsetto</i>
<i>Schmuck</i>	8	Repetition / absolute universalisation	<i>Schmuck / imbranato / fesso / babbeo</i>
<i>Yenta</i>	2	Repetition / absolute universalisation	<i>Yenta / ultima ruota del carro</i>
<i>Schnorrer</i>	3	Repetition / universalisation	<i>Schnorrer / ci fa la carità</i>
<i>Shiksa</i>	2	Repetition	<i>Shiksa</i>
<i>Goy</i>	15	Linguistic translation / repetition	<i>Non ebreo / goy</i>
<i>Chutzpah</i>	1	Repetition	<i>Chutzpah</i>
<i>Oy vey</i>	1	Deletion	-
<i>Mazel tov</i>	7	Repetition / deletion	<i>Mazel tov</i>
<i>Kvetch</i>	2	Absolute universalisation	<i>Lamentela / lamentarsi</i>
<i>Kvell</i>	2	Repetition / absolute universalisation	<i>Kvell / molto felici</i>
<i>Bupkis</i>	1	Absolute universalisation	<i>Non fa ridere</i>

Translation strategies for *Yiddish realia* in *The Marvelous Mrs. Maisel*.

The present chart shows clear evidence of four main translation strategies taken from Aixelà's (1996) categorisation. The dominant strategy adopted is undoubtedly repetition, notably with ethnographic and religious realia, transferred in the target text as they are in the source text. When dealing with terms generally well-known and having a consolidated equivalent in the Italian vocabulary, the translator opts for a linguistic translation (e.g. *rabbi* as *rabbino*). For realia describing socio-political situations or states of mind, translators rely on strategies of absolute or limited universalisation. The Italian language lacks cultural equivalents, therefore it is difficult to maintain these terms without generating a level of opacity or losing their cultural boundedness. However, as their force depends on a connection between cultural specificity and the effect on the target audience, by using universalisation strategies translators opt for more neutral solutions that can still preserve the general effect produced on the receiver. Lastly, when a Yiddish term is considered difficult to render in the target text and its cultural specificity appears to be less relevant in context, subtitlers can also choose to omit the realia using a deletion strategy.

3.6.1 Yiddish Realia: a Qualitative Analysis of Selected Cases

While the overall pattern is clear, a number of individual cases deserve closer qualitative attention as they illustrate more complex translation challenges. Where relevant, the official Amazon Prime Video subtitle rendition has been manually transcribed and included for comparison. This comparative approach has revealed interesting divergences between fansubs and official subtitles.

Following the order presented in the chart above, the first case to examine is *mitzvah*, which literally means “commandment” and indicates a meritorious act that expresses God’s will (Rosten, 1968, p. 255 *mitzvot*). In most instances in the corpus, *mitzvah* is rendered with a repetition when referring to the name of the bar in the series, as proper names are usually not subject to translation. However, when used as a religious term with its religious signification, absolute universalisation can also be used, as in the following instance:

ST: It’s a big *mitzvah* to make a shidduch.

TT: È il *precetto* per potersi fidanzare.

(*The Marvelous Mrs. Maisel*, S02E04, 29:15)

Fansubs and official subtitles propose the same rendition, with *precetto* as the Italian equivalent chosen to render the realia. *Precetto*, however, is a more general term in Italian that indicates a

“principle, norm or mandatory rule of conduct” (Treccani, n.d., voce *precetto*, my translation), therefore, the cultural specificity of the Jewish term is inevitably lost in translation. Furthermore, the term *shidduch*, indicating an arranged marriage (Rosten, 1968, p. 347 *shiddach*), is also generalised into a neutral *fidanzarsi*, completely omitting the Yiddish specificity, as well as the semantic notion of an arranged match. Consulting the OpenSubtitles reference corpus confirms that *mitzvah* is almost always transcribed in the target text without translation. A more faithful rendition might have been, in this case, to preserve both realia, relying on the context to convey their culture-bound meaning to the Italian audience.

Another interesting case to analyse is that of *Hannukah* or *Hanukkah* or *Chanukkah*²³ also known as the Feast of Lights, an eight-day celebration in December (Rosten, 1968, p. 71, *Chanukah*). In the chart it is shown that the word is mostly repeated as it is, in two orthographic forms, *Hanukkah* (most common) and *Chanukkah* (a transliteration of Hebrew). In one instance, however, the realia is omitted:

ST: I did a version of the *Hanukkah* run a couple weeks ago.

TT: Avevo fatto lo sketch due settimane fa.

OS²⁴: L’avevo fatto anche qualche settimana fa.

(The Marvelous Mrs. Maisel, S01E06, 17:01)

This case is worth mentioning, as it is a perfect example of what Florin (1993) calls the importance in context of the realia, one of the considerations to make when choosing the translation strategy to adopt. In this case, in fact, Midge uses *Hannukah* as an attribute to identify the sketch she just made to her manager Susie, who already knows which one she is referring to. Furthermore, the sketch had just been shown in the previous scene, therefore the realia could be easily omitted without causing any particular cultural loss. Indeed, this same translation choice was made by both the fansubbers and the official subtitler.

The rendition of the realia *mohel* introduces two interesting cases to observe, despite its limited occurrences. A *mohel* is a ritual circumciser (Rosten, 1968, p. 259, *mohel*), a crucial figure in the Jewish culture.

²³ This is one of the cases in which the Yiddish term can be written in different ways, therefore, to make sure all instances of the lemma were shown, when querying the corpus it was necessary to type the search as follows: [word="hannukah|hanukkah|chanukah|Hannukah|Hanukkah|Chanukah"]. The same pattern was used for *Aliyah*, *schmooze*, *schmuck*, *goy*, *chutzpah*, *kvetch* and *kvell*.

²⁴ OS = Official subtitles from Amazon Prime Video. When the alternative is not present, it is implied that the version used by the official subtitler is the same as the one produced by the fansubbers.

ST: Every cut of the *mohel's* knife on vivid display.

TT: Così tanto da mettere in bella vista la *circoncisione*.

(*The Marvelous Mrs. Maisel*, S02E06, 05:51)

This is a case of deletion and absolute universalisation together, as the action undertaken by the *mohel* in the source text is expressed with *cut of the mohel's knife*, and in the target version it is simply rendered as *circoncisione*. The culturally specific item is omitted and substituted by a more general and neutral concept, that can be easily accepted and understood by the target audience, as it is widely known that Jewish people undergo this procedure. It is, however, a relevant form of cultural neutralisation, as a *mohel* is not just a doctor, but a religious figure all-round.

ST: The doctor is a man called a *mohel*, who has no medical training, nope.

TT: E il dottore è un tale chiamato *mohel* che non ha alcuna formazione medica.

(*The Marvelous Mrs. Maisel*, S03E07, 33:08)

This instance is also interesting as the rendition does not exactly replicate the irony of the original text. The original scene is entirely based on a succession of sarcastic remarks questioning the competence of the *mohel*. Apparently, this can pass as a repetition that preserves the cultural specificity of the realia. However, using the expression *un tale chiamato mohel* it almost seems as if that is the actual name of that doctor: the reader might interpret it as the proper name, not as a religious denomination. To avoid this misunderstanding, a better solution might have been “E il dottore è il cosiddetto *mohel*”.

Culinary realia offer a series of remarks on the translation strategies adopted as well. It is the case, for example, of *kugel*, “a pudding of noodles or potatoes” (Rosten, 1968, p. 201, *kugel*), which has been rendered with several translation strategies. Other than repetition, in some cases it has also been rendered through a limited universalisation with the word *pudding*; i.e. a word belonging to the same culinary field, but without the Jewish specificity of the original. Other possible renditions found in the OpenSubtitles reference corpus are: *sformato*, *timballino*. Another form of absolute universalisation is carried out with the food realia *latke*, a potato pancake (Rosten, 1968, p. 210, *latke*), in the following occurrence:

ST: Tonight's vodka and *latke* mixer will be in the rec room at 8 p.m..

TT: Stasera troverete i *cocktail* nella sala ritrovo alle 20.00.

(*The Marvelous Mrs. Maisel*, S02E04, 20:49)

The realia here is omitted and reduced to a general *cocktail*. Not only is there a loss in terms of cultural specificity, but also in terms of meaning as the food is not mentioned at all and the specific drink *vodka* is generalised. Instances of the word appear in the reference corpus, which shows that *latke* could have been translated as *frittata*, *frittella*, *frittella di patate*, *pancake*, *dolcetti*, *tortel di patate*. These are interesting occurrences, however, the most suitable might have been *frittata* or *frittata di patate*, which contains the actual ingredients of the recipe, even though there is still a loss in culture specificity. One of the instances found in the reference corpus, *tortel di patate*, a traditional dish typical of the Trentino region in Italy, represents an attempt at naturalisation (Aixelà, 1996), substituting the Jewish culinary realia with an Italian regional equivalent. While this strategy preserves the specificity of the traditional dish, it inappropriately displaces the cultural reference from a strongly culture- and religious-bound recipe, to a typical Italian dish, which would also appear completely out of context in the Ashkenazi community setting.

When looking at the translation of the word *schlep*, another interesting case of absolute universalisation can be observed. The verb *schlep* means to drag, pull or lag behind (Rosten, 1968, p. 354, *schlep*) and in the corpus it is rendered as a neutral *visitare*.

ST: And there's no need to *schlep* to the Museum of Natural History.

TT: Non c'è bisogno di *visitare* il Museo di Storia Naturale.

(*The Marvelous Mrs. Maisel*, S04E06, 51:22)

To *schlep* implies an activity that is tiring, whereas in this instance *visitare* gives the action a pleasant connotation, opposite to the original meaning. Better renditions of the verb can be found in the reference corpus, such as *trascinare*, *trascinare avanti e dietro*, *scarrozzare*. Among these, “trascinarsi fino al Museo di Storia Naturale” would have been the most faithful one, preserving the physical effort implied in the realia.

The word *schmuck* can be used to indicate an ornament, the male genitalia or a detestable fellow (Rosten, 1968, p. 365, *shmuck*). In the series, it is mostly present with the last interpretation and is rendered with an absolute universalisation strategy as *fesso*, *babbeo*, *scemi* or *imbranato*. However, some interesting observations deserve attention.

ST: I'm Jewish, *shmuck*.

TT: Sono ebreo, *scemi*.

OS: Sono ebreo, *shmuck*.

(*The Marvelous Mrs. Maisel*, S01E08, 04:52)

This is a case of divergence in the two renditions made by fansubber and official subtitler. The former adopts an absolute universalisation strategy, the latter adopts a repetition strategy. The official subtitler, here, chooses a more faithful rendition as in this context it is important to keep the Yiddish term as a reaffirmation of the Jewish nature of the protagonist. Translating it with an Italian equivalent such as *scemi* does not render the same polemical effect.

ST: The *shmuck* playing cards in the storeroom?

TT: L'*imbranato* che gioca a carte nel deposito?

(*The Marvelous Mrs. Maisel*, S02E02, 40:25)

Among the different Italian equivalents used, attention should be paid to *imbranato* as it does not actually preserve the same offensive characterisation as the realia, rather it is much more neutral. It is extremely relevant here to query the reference corpus in search of more accurate renditions. Among the several alternatives found in the corpus, *sfigato* or *cretino* appear to be the most faithful renditions, as they preserve the offensive connotation without resorting to stronger Italian swearwords²⁵ that would be excessive.

The following example of absolute universalisation of *yenta* is extremely interesting. The word indicates a woman of low origins or a gossip woman (Rosten, 1968, p. 436, *yenta*).

ST: The chopped liver *yenta* from Allerton Avenue?

TT: L'*ultima ruota del carro* di Allerton Avenue?

(*The Marvelous Mrs. Maisel*, S05E09, 04:03)

The Italian equivalent chosen, *l'ultima ruota del carro*, can indicate the same idea of being considered of low importance, but loses the gossip nuance of meaning, in addition to the loss of the cultural specificity as the typical Ashkenazi woman is replaced by a generalisation. Moreover, the use of *chopped liver* is equally remarkable. The idiomatic English expression

²⁵ Other renditions present in the reference corpus are: stronzi, idiota, scarto, sciocco, coglione, figlio di buona donna, bastardo, schifoso, stupido.

being chopped liver means “being irrelevant”; this nuance is captured in the target rendition, but the translator loses the culinary realia, as chopped liver is a typical dish of the kosher cuisine. Taking into consideration other alternatives given by the reference corpus, a more faithful rendition might have been *ficcanaso* or *pettegola*, which preserve the gossipy connotation of the original. However, in these alternatives the low origin dimension disappears. It is a case of cultural asymmetry for which Italian does not have a term that can capture both dimensions simultaneously and every translation inevitably entails a degree of loss due to cultural asymmetry (Aixelà, 1996).

A similar situation happens with *kvell*, meaning to gush, to be proudly happy (Rosten, 1968, p. 203, *kvell*).

ST: Your parents must be *kveling*.

TT: I tuoi genitori *saranno molto felici*.

(*The Marvelous Mrs. Maisel*, S02E05, 29:45)

In the present occurrence, the equivalent chosen is much more neutral. Italian lacks, as it was for *yenta*’s case, a single term that captures the intensity and cultural connotation of the realia, making some degree of loss unpreventable. The reference corpus does not provide a useful alternative as it only offers 2 instances, but a possible solution could have been simply *I tuoi genitori staranno scoppiando d’orgoglio*.

A significant shift of perspective in translation occurs in the case of the absolute universalisation used for *schnorrer*, a noun indicating a beggar, a mendicant (Rosten, 1968, p. 369, *schnorrer*). The following Italian rendition is a peculiar case as it switches the point of view:

ST: Your mother thinks we’re *schnorrers*.

TT: Tua madre *ci fa la carità*.

(*The Marvelous Mrs. Maisel*, S04E02, 11:42)

Although the overall sense is maintained and despite the inevitable cultural loss, a more accurate alternative exists in the reference corpus: *scrocconi*. *Tua madre ci considera degli scrocconi* would have been more faithful, as it preserves at least the original subject.

One last instance to observe is *bupkis*, which is used to refer to something trivial, worthless, non-sensical (Rosten, 1968, p. 55, *bupkis*).

ST: Back in America that line gets *bupkis*.

TT: In America questa battuta *non fa ridere*.

(*The Marvelous Mrs. Maisel*, S02E01, 37:11)

The Italian rendition *non fa ridere* is another absolute universalisation that replaces the Yiddish realia with a neutral equivalent that communicates the general meaning but loses the specificity. Among the alternatives found in the reference corpus, one that appears to be the most adequate in context is *non vale una cicca*, as it preserves the colloquial register and the connotation of worthlessness of the original realia. A possible translation, therefore, could have been: *In America quella battuta non vale una cicca*.

3.6.2 Yiddish Realia: Concluding Remarks

From the present analysis of culture-specific terms as marked discourse in *The Marvelous Mrs. Maisel* dominant general tendencies have emerged. After classifying realia in subcategories in order to organise their analysis (geographic, ethnographic and socio-political realia), according to the literature presented in Chapter 1, it becomes apparent that each group privileges specific translation strategies over others, and implies a series of translation issues to address.

The dominant strategies that emerged from the analysis have been repetition, mainly for religious and ethnographic realia, and absolute universalisation (or functional equivalence) for socio-political realia and states of mind that do not present a direct equivalent in Italian. This divergence might be justified by a different degree of cultural accessibility of these categories, in compliance with Florin's criteria (1993) discussed in Section 1.6: ethnographic items, being part of the identity of a culture, are more prone to exposure to the general public through mass media. Conversely, socio-political realia involve institutional systems and their preservation in the target environment might render the rendition more opaque; hence, the translator resorts to absolute universalisation to ensure accessibility sacrificing the cultural reference.

Differently from the analysis and rendition of other kinds of marked discourse, such as slang and profanities examined in the case of *Fleabag*, when dealing with culture-bound terms, a partial or total loss in translation is nearly always inevitable. As Aixelà (1996) observes, cultural asymmetry between linguistic communities inevitably produces a level of opacity in the target culture and language that no translation strategy can fully compensate.

Finally, the contrastive approach between fansubs and official subtitles, in this case does not reveal a general pattern of strategies adopted. Indeed, in most instances the two versions

coincide perfectly, perhaps because realia items tend to be more culturally bound, leaving less room for creative recreations. Only in rare cases, as it was shown for the case of *schmuck*, the official subtitles appear to be more faithful than the fansubbed version.

CONCLUSIONS

The primary aim of this research was to investigate challenges in rendering subtitles in two contemporary audiovisual products characterised by marked linguistic features. It is precisely this dialogic dimension, of a text written to sound spoken that makes the subtitling of a TV series particularly challenging, requiring a crucial focus on the rendition of conversational markers. By associating theoretical Audiovisual Translation (AVT) studies with Corpus Linguistics methodologies, this thesis presents a systematic primarily quantitative framework of analysis for the two designated TV series (*The Marvelous Mrs. Maisel* and *Fleabag*). The compilation of specialised frequency lists and keyword analysis for the two audiovisual products successfully confirmed the presence of the core features assumed by intuitive observation of both comedies. Lexical density and speech fragmentation, taboo language and idiomatic slang expressions, together with culture-specific terms (realia) were favourably encountered as distinctive linguistic traits and empirically evaluated through the present analysis.

First and foremost, the quantitative evaluation demonstrated a consistent pattern of text reduction across both corpora of data. In coherence with the simplification hypothesis (Baker, 1993) and condensation tendencies commonly encountered in AVT (Pavesi, 2019) a systematic token reduction from source to target text was observed, ranging from 15 to 17% in both series. However, a closer look at the data revealed a higher reduction rate in *The Marvelous Mrs. Maisel*, supporting the initial expectation of a lexically more dense script that would consequently pressure the subtitler for a stronger use of condensation and reformulation strategies in target renditions. Considering AVT's literature, the reasons behind this total or partial reduction (Díaz-Cintas and Remael, 2007) have been explored in the omission of discourse markers as the primary markers of speech usually subject to omission when faced with technical restrictions and translational challenges. The three cases analysed (*well, you know, I mean*) confirm omission and reformulation as the two main strategies adopted, with a coherent higher reduction rate in *The Marvelous Mrs. Maisel*.

More specifically, in *The Marvelous Mrs. Maisel* long nominal phrases were frequently neutralised through lexical condensation and resorting to more neutral reformulations in Italian. While these solutions ensured easier readability and accessibility in compliance with the technicalities of the audiovisual medium with its spatial and temporal limitations, the distinctive fast-talking and nonsense rambling that made Midge Maisel's character recognisable is inevitably attenuated. Her original voice is neutralised, confirming Arias-Badia's (2015)

struggle with the neutralisation of the creative voice of the characters. Conversely, the research shows that *Fleabag*'s characteristic fragmented speech is generally stylistically preserved, as demonstrated in the analysis of punctuation marks and their technical grammar in the genre. In fact, while dashes have not proved to be reliable sources of fragmentation, as they mostly serve technical purposes, suspension points confirm to be productive indicators of how hesitations and interrupted utterances are maintained throughout the series.

Another counter-tendency in the analysis of *Fleabag* emerged in the management of swearwords and profane vocabulary in the Italian renditions. Indeed, contrary to the traditional AVT neutralisation hypothesis, the data shows an 84% preservation rate of taboo language rendered through direct equivalents in the target text. Once again, subtitlers recognise the perlocutionary force and the role played by these marked expressions in the characterisation of the protagonists and make a greater effort to successfully adapt these expressions, so as to not impact the ironic nuances of the series.

Marked language also involves idiomatic and slang expressions, that here undergo the same kind of process: the dominant strategy emerged from the examination is substitution (Díaz-Cintas and Remael, 2007), even though the manual annotation and corpus-assisted analysis have shown a variety of strategies based on the context and the data under examination. In fact, the contrastive analysis between the fan-produced subtitles and the professional ones has shown how institutional constraints in the AVT field influence the work of translators. Fansubs exhibit a general predisposition towards more creative, source-oriented and foreignising solutions, preserving the otherness of the original text; whereas, official subtitles are more inclined towards neutral, target-oriented and domesticating approaches, and therefore more in line with the expectations of AVT literature.

A similar phenomenon was observed in the case of realia in *The Marvelous Mrs. Maisel*, where translation strategies differed depending on the semiotic category of culture-specific item, with ethnographic and religious realia showing a tendency towards repetition and socio-political items resorting to absolute universalisation (Aixelá, 1996). In this instance, interestingly, the contrastive analysis between fansubs and original subtitles did not exhibit opposite tendencies, but rather coincided in the vast majority of cases. This result suggests that culture-bound items leave little room for creative recreations, forcing both amateur and professional translators to remain faithful to the original utterance. Moreover, it reasserts how a certain level of culture asymmetry between source and target text remains impossible to bridge.

While this research generally achieved its analytical goals, several limitations must be acknowledged, which open up productive discussions for future inquiry. First and foremost, this study privileged quantitative analysis, whereas qualitative examinations were restricted to a targeted selection of representative examples that confirmed the expected results. As a consequence, it is implied that expanding the research to more elements such as other discourse markers, a wider range of profanities or varied distinctive markers of orality would undoubtedly strengthen the systematicity of the results. Moreover, the research attempted a partial contrastive analysis between the two picked audiovisual products, and continued with a targeted examination of distinctive features of each series. However, as both series belong to the comedic genre, interesting findings might have emerged analysing more humorous elements in both series.

Another limitation lies in the textual nature of the corpus data at hand, since subtitles work better when interacting with the visual and acoustic channels of the audiovisual medium. Consequently, the possibility of future research through a multimodal corpus that incorporates image and sound would ensure ample room for developments in examining the use of compensative strategies for omissions, moving beyond a lemma-level analysis and towards an utterance-level context. In fact, a purely quantitative calculation of the omission rate represents a starting point to identify the systematic nature of the translation choices of the subtitlers. For instance, as observed in the analysis above, the absence of direct equivalents does not always imply omission from the translator, but rather an asymmetry between English and Italian. Simple lemma queries do not account for compensatory additions in the text: for example, the subtitler might have introduced a slang term or profanity not present in the original text to compensate for a possible deletion in order to maintain the overall register.

Another starting point of this study is given by the selection of the dataset: two contemporary female-led comedies with similar yet distinct features at the same time can only provide systematic findings to a certain extent. Therefore, it would be interesting to achieve a more exhaustive overview of this television genre: future research could expand the hypothesis generated here by intuitive observation and employ this methodological model to a wider catalogue of series, to evaluate if the tension between text reduction and stylistic preservation of orality is persistent across the genre.

A final methodological choice that undoubtedly influenced the outcomes of the present study concerns the construction of the corpora using fan-generated subtitles. Fansubs admittedly represent productive environment for analysing the creativity of the subtitler and, despite their amateur nature, demonstrate an overall professionalism, aligning with the original

version in the majority of the cases examined here. Furthermore, their non-commercial character gives them the freedom to somewhat operate outside institutional and commercial constraints. Had the study been conducted using exclusively official subtitles, which remain inaccessible for corpus compiling due to strict copyright norms of the streaming platforms, the empirical results might have reflected a stronger adherence to the systemic neutralisation of the character's voices extensively inquired in AVT studies. This acknowledgment is essential to highlight how the identified translation tendencies are strongly biased by the environment in which the subtitles are produced, especially when dealing with mainstream products, which have been subject to a wide variety of renditions.

In conclusion, the tension between text reduction due to medium constraints and stylistic preservation remains open to further exploration. This study provides a replicable methodological model for the systematic analysis of fictional orality in comedy subtitling, revealing that neutralisation is not a universal tendency but a context-dependent phenomenon shaped by the linguistic features of the series and of the marked discourse under examination, as well as the environment in which the translation is produced. The contrasting and varied strategies emerged from this investigation have demonstrated how subtitling cannot be considered a simple case of mechanical and systematic condensation, but rather an intricate web of linguistic negotiations.

BIBLIOGRAPHY

- Aixelá, Javier Franco (1996). 'Culture-specific Items in Translation.' In Álvarez & Vidal (eds.), *Translation, Power, Subversion*. Clevedon: Multilingual Matters, pp. 52–78.
- Arias-Badia, B. (2015). Towards a methodology for the analysis of neutralisation in Spanish subtitling. In G. Corpas, M. Seghiri, R. Gutiérrez, & M. Urbano (Eds.), *New horizons in translation and interpreting studies: Proceedings of the 7th AIETI conference* (pp. 513–526). Tradulex.
- Arias-Badia, B. (2020). *Subtitling television series: A corpus-driven study of police procedurals*. Peter Lang.
- Arias-Badia, B. (2021). Using corpus pattern analysis for the study of audiovisual translation: A case study to illustrate advantages and limitations. In M. Calzada & S. Laviosa (Eds.), *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection* (MonTI, 13, pp. 93–113).
- Attardo, S. (1994). Introduction. In *Linguistic theories of humor* (pp. 2–13). Mouton de Gruyter.
- Attardo, S. (2017). The general theory of verbal humor. In S. Attardo (Ed.), *The Routledge handbook of language and humor* (pp. 126–142). Routledge.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). John Benjamins.
- Baker, P. (2020). Corpus-assisted discourse analysis. In C. Hart (Ed.), *Researching discourse: A student guide* (pp. 124–142). Routledge.
- Baños-Piñero, R., & Chaume, F. (2009). Prefabricated orality: A challenge in audiovisual translation. inTRAlinea. Special Issue: *The Translation of Dialects in Multimedia*.
- Beaumont, J. (2021). Phoebe Waller-Bridge's *Fleabag(s)*: Direct address and narrative control from stage to small screen. *Journal of International Women's Studies*, 22(2), 103–119.
- Chaume, F. (2004a). Discourse markers in audiovisual translating. *Meta*, 49(4), 843–855. <https://doi.org/10.7202/009785ar>
- Chaume, F. (2004b). Film Studies and Translation Studies: Two Disciplines at Stake in Audiovisual Translation. *Meta*, 49(1), 12–24. <https://doi.org/10.7202/009016ar>
- Chaume, F. (2013). 'The turn of audiovisual translation: New audiences and new technologies.' *Translation Spaces*, 2(1), pp. 105–123.
- Díaz Cintas, J. (2006). Fansubs: Audiovisual translation in an amateur environment. In J. Díaz Cintas (Ed.), *Audiovisual translation: Subtitling* (pp. 37–52). St. Jerome Publishing.
- Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St Jerome.

Díaz-Cintas, J. (2008). Teaching and learning to subtitle in an academic environment. In J. Díaz-Cintas (Ed.), *The didactics of audiovisual translation* (pp. 89–104). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.77.10dia>

Dore, M. (2020). *Humour in audiovisual translation: Theories and applications*. Routledge.

Dore, M. (2024). Scripted stand-up comedy in translation: The Marvellous Mrs Maisel dubbed in Italian. *Traduction et Langues*, 23(2), 96–117.

Florin, Sider (1993). 'Realia in Translation.' In Zlateva, Palma (ed.), *Translation as Social Action: Russian and Bulgarian Perspectives*. London: Routledge, pp. 122–128.

Gambier, Y. (2003). Introduction: Screen Transadaptation: Perception and Reception. *The Translator*, 9(2), 171–189. <https://doi.org/10.1080/13556509.2003.10799152>

Gottlieb, H. (1992). Subtitling: A new university discipline. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: Training, talent and experience*. Papers from the First Language International Conference, Elsinore, Denmark, 1991 (pp. 161–170). John Benjamins. <https://doi.org/10.1075/z.56.26got>

Gottlieb, H. (1994). Subtitling: Diagonal translation. *Perspectives: Studies in Translatology*, 2(1), 101–121. <https://doi.org/10.1080/0907676X.1994.9961227>

Hanks, P. (2012). How people use words to make meanings: Semantic types meet valencies. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 52–67). Masaryk University Press.

Istituto della Enciclopedia Italiana. (n.d.). *Cinghia*. In *Treccani online dictionary*. Retrieved May 7, 2026, from <https://www.treccani.it/vocabolario/cinghia/>

Ivarsson, J., & Carroll, M. (1998). *Subtitling*. TransEdit.

Johansson, Stig (2007). *Seeing Through Multilingual Corpora*. Amsterdam: John Benjamins.

Kennedy, Graeme (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

Laviosa, S., & Liu, K. (2021). The pervasiveness of corpora in translation studies. *翻譯季刊 (Translation Quarterly)*, 101, 5–20.

Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 133–149). Brill. https://doi.org/10.1163/9789401203791_009

Lexical Computing CZ s.r.o. (2024). *Sketch Engine* [Computer software]. <https://www.sketchengine.eu/>

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Myhre, N. (2024). *Subtitle Edit* (Version 4.0.5) [Computer software]. <https://www.nikse.dk/subtitleedit/>

Olohan, M. (2004). *Introducing corpora in translation studies*. Routledge.

Oxford University Press. (n.d.). *Cheap thrill*, n. In *Oxford English dictionary*. Retrieved May 7, 2026, from <https://doi.org/10.1093/OED/8119881160>

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins.

Pavesi, M. (2009). Dubbing English into Italian: A closer look at the translation of spoken language. In J. Díaz Cintas (Ed.), *New trends in audiovisual translation* (pp. 197–209). Multilingual Matters. <https://doi.org/10.21832/9781847691552-016>

Pavesi, M. (2019). Corpus-based audiovisual translation studies: Ample room for development. In L. Pérez-González (Ed.), *The Routledge handbook of audiovisual translation* (pp. 315–333). Routledge.

Pedersen, J. (2010). "When do you go for benevolent intervention? How subtitlers determine the need for cultural mediation". In *New Insights into Audiovisual Translation and Media Accessibility*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789042031814_007

Peromingo, J. P. R., Martín, R. A., & Riaza, B. G. (2014). New approaches to audiovisual translation: The usefulness of corpus-based studies for the teaching of dubbing and subtitling. In E. Bárcena, T. Read, & J. Arús (Eds.), *Languages for specific purposes in the digital era (Vol. 19)*, pp. 303–322). Springer.

Ranzato, I. (2010). "Localising Cockney: translating dialect into Italian". In *New Insights into Audiovisual Translation and Media Accessibility*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789042031814_010

Romero-Fresco, P. (2009). Naturalness in the Spanish dubbing language: A case of not-so-close friends. *Meta: Journal des traducteurs*, 54(1), 49–72. <https://doi.org/10.7202/029793ar>

Rollo, A. (2014). Humour e giochi di parole in “Astérix et Obélix: Mission Cléopâtre”. Quali strategie nella traduzione audiovisiva? In G. L. De Rosa, F. Bianchi, A. De Laurentiis, & E. Perego (Eds.), *Translating humour in audiovisual texts* (pp. 129–154). Peter Lang.

Rosten, Leo (1968). *The Joys of Yiddish*. New York: McGraw-Hill.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse* (R. Carter, Ed.). Routledge.

- Soffritti, M. (2019). Multimodal corpora in audiovisual translation studies. In L. Pérez-González (Ed.), *The Routledge handbook of audiovisual translation* (1st ed., pp. 334–349). Routledge. <https://doi.org/10.4324/9781315717166>
- Steinmetz, Sol (1986). *Yiddish and English: A Century of Yiddish in America*. Tuscaloosa: University of Alabama Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.
- Toury, G. (2012). *Descriptive translation studies and beyond* (Revised ed.). John Benjamins.
- Treccani. (n.d.). [*Precetto*]. Enciclopedia Treccani. URL
- Veiga, M. J. (2009). The translation of audiovisual humour in just a few words. In J. Díaz Cintas (Ed.), *New trends in audiovisual translation* (pp. 158–175). Multilingual Matters. <https://doi.org/10.21832/9781847691552-014>
- Zabalbeascoa, P. (2005). Humor and translation—An interdisciplinary. *Humor: International Journal of Humor Research*, 18(2), 185–207. <https://doi.org/10.1515/humr.2005.18.2.185>
- Zabalbeascoa, P. (2008). The nature of the audiovisual text and its parameters. In J. Díaz-Cintas (Ed.), *The didactics of audiovisual translation* (pp. 21–37). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.77.05zab>
- Zabalbeascoa, P., & Hayes, L. (2024). AVT modes and processes. In P. Zabalbeascoa (Ed.), *Audiovisual translation* (Chap. 4, pp. 68–85). Routledge.
- Zanettin, F. (2012). *Translation-driven corpora*. Routledge.

ONLINE REFERENCES

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>

Jakubíček, M., Kilgariff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July). *The TenTen corpus family*. In 7th International Corpus Linguistics Conference CL (pp. 125-127).