



UNIVERSIDAD DE JAÉN
Facultad de Humanidades y Ciencias de la Educación

Trabajo Fin de Grado

**On the translation of
taboo words in an
English-Spanish corpus
of film subtitles**

Alumno/a: Antonio Torres Cuenca

Tutor/a: Prof. D. Francisco Javier Díaz Pérez
Dpto.: Filología inglesa

Junio, 2016

Table of contents

Index of figures, tables and graphs.....	ii
Abstract.....	iii
1. Introduction.....	1
2. Audiovisual Translation.....	2
3. Linguistic corpora.....	6
4. Pragmatic functions and translation of taboo words.....	12
4.1. Pragmatic functions of taboo words.....	12
4.1.1. Pragmatic functions of ‘fuck’ and its derivative ‘fucking’.....	14
4.1.2. Pragmatic functions of ‘shit’.....	15
4.2. Translation of taboo words in subtitling.....	16
5. Methodology.....	19
5.1. Objectives.....	19
5.2. Research instrument.....	19
5.3. Research stages.....	21
6. Results.....	23
6.1. General strategies for translating taboo words.....	23
6.1.1. General translation strategies used for translating ‘fuck’ and ‘fucking’.....	25
6.1.2. General translation strategies used for translating ‘shit’.....	27
6.2. Variables of the translation of taboo words.....	28
6.2.1. Grammatical category.....	29
6.2.1.1.Fuck and fucking.....	29
6.2.1.2.Shit.....	33
7. Conclusion.....	36
Bibliographical references.....	38

Index of figures, tables and graphs

Figure 1. Search of the word “fuck” in the <i>Veiga Corpus</i>	10
Figure 2. Example of the English word “fuck” in context.....	10
Table 1. General uses of the taboo word ‘fuck’ and its derived words.....	14
Table 2. General uses of the taboo ‘shit’.....	15
Figure 3. Word searched for together with its context.....	20
Figure 4. How to select the context where the searched word is wanted to be shown.....	20
Figure 5. Word searched for together with its context.....	20
Figure 6. Two examples of a search in <i>OpenSubtitles</i> using CQP.....	21
Figure 7. Examples extracted from the Excel spreadsheet.....	22
Table 3. Translation strategies found in the corpus with a representative example.....	24
Graph 1. Translation strategies of taboo words identified within <i>OpenSubtitles</i>	25
Graph 2. Translation strategies used for translating ‘fuck’ and ‘fucking’.....	25
Table 4. General translation strategies used to translate ‘fuck’ and ‘fucking’.....	26
Graph 3. Frequency of use of translation strategies for translating ‘shit’.....	27
Table 5. General translation strategies used to translate ‘shit’.....	28
Table 6. Frequency of grammatical categories of the word ‘fuck’.....	29
Table 7. Translation strategies used in ‘fuck’ and ‘fucking’.....	30
Table 8. Frequency on which grammatical categories of ‘shit’.....	33
Table 9. Translation strategies used in ‘shit’ considering the grammatical class.....	33

Abstract

The main objective of the present paper is to analyse the different translation strategies used to translate taboo words from English to Spanish in film subtitles. With this purpose in mind, a computerized multilingual corpus of film subtitles, *OpenSubtitles*, has been used to extract empirical data of English as source language and Spanish as target language. After a theoretical review, the paper focuses on the methodological aspects of the study as well as the obtained results. With regard to the latter, even though the commonest translation strategy used to render these words is literal translation, in a high number of cases the taboo word has been omitted or softened and replaced by a neutral word. After analysing all the samples, it has been also determined that the grammatical category of the taboo word influences the way in which it is translated.

Key words: Audiovisual translation, Subtitling, Taboo words, Corpus Linguistics, films.

1. Introduction

This End of Bachelor Degree Project is aimed at carrying out initial research in Audiovisual Translation (AVT) and based on linguistic corpora, which according to Baños, Bruti and Zanotti (2013: 483), is “a methodology for language studies that has been successfully applied to translation studies in general”. Linguistic corpora offer large-scale empirical samples of language in use. These samples may involve either spoken or written language instances. Apart from monolingual corpora, constructed by samples of just one language, there are also bilingual and multilingual corpora, respectively involving two languages and more than two different languages.

In this paper, the *OpenSubtitles* corpus will be the object of study and the source of all the examples that will be used in the present study. This research tool offers a compilation of film subtitles from many different languages organized in an aligned way. The possibilities offered by this corpus make it an excellent tool for research in the field of audiovisual translation. In this study, English has been the source language and Spanish language has been the target language, or language translated into.

First of all, a brief theoretical introduction to the field of audiovisual translation is presented in chapter 2. Chapter 3 focuses its attention on corpora and its several utilities within the audiovisual translation field. In the following chapter, the different pragmatic function of swearwords as well as the translation of this type of lexicon will be considered. This will be followed by the methodology chapter. The different results extracted from this research will be illustrated afterwards, followed by the conclusions drawn from this study. In this sense, it is important to highlight that each and every example shown in this essay has been extracted from *OpenSubtitles*.

This paper is aimed at analysing the translation strategies of the most frequent English swearwords, ‘fuck’ and its derived word ‘fucking’ and ‘shit’. The main reason why these words are going to be scrutinized is because of their high frequency in current everyday speech. As mentioned above, these are the most used taboo words in English, and due to the pragmatic functions they fulfil in everyday speech and in film dialogues, as an intended reflection of natural speech.

2. Audiovisual Translation

Nowadays, owing to the many different cultures and languages that exist around the world together with the growth of international communication, translation is a phenomenon that occurs every day. It is obvious that languages are closely connected with culture. In fact, language is one of the ways whereby culture is expressed. That is the reason why Ponce Márquez (2007) said that translators are mediators between two different linguistic and cultural systems: on the one hand, the source text (ST) addresser deals with the system of the source language (SL) or the language translated from; on the other hand, the source text (ST) addressee deals with the system and culture of the target language (TL) or the language translated into which the ST is translated. Consequently, translators connect different cultures trying to cope with both the differences and incompatibilities of those cultures. Some examples of such differences regard different ideologies or the consideration of some words as taboo concepts. Translators are transmitters as well as receivers of a message. Hence, the proficient translator must be able to think as a member of both source and target culture, as well as to master both linguistic systems in order to sound natural to the TL addressee. In other words, if translators paid attention only to linguistic matters and not to the extra-linguistic components of the message, they would not be able to show the social and situational components of the ST in the TL. Consequently, the task of producing an accurate translation resides mainly in the translator; hence each professional may render the same text in a different way. The reason that can explain this fact is that translation is a diverse and heterogeneous process rather than a homogeneous one. Nonetheless, even though each person's contribution may be different, translators must look for *adequacy* between both languages and *acceptability* concerning the audience of the TL. The first criterion implies that an accurate translation must preserve the communication effectiveness of the original text in the TT. Acceptability, in turn, refers to the way in which a native speaker of the TL perceives the translation. If it sounds natural, it will not breach the acceptability principle. To sum up this idea, when translating, not only linguistic issues will have to be addressed, but also ideological, cultural, and many other aspects concerning the differences between the SL and the TL.

Even though there are many different types of translation, this paper focuses only on audiovisual translation (AVT), which encapsulates the translation of any audiovisual text where an interaction between sound and images takes place. The

sending of the audiovisual message is performed thanks to the visual and the acoustic channels. For this reason, as Pettit (2009: 44) asserts, “an audiovisual text offers a cultural representation of the world, both through language and the image”. In the field of AVT, translators tackle translations of TV series, films, TV programmes, videogames, and so on and so forth. Thus, they must consider verbal and nonverbal aspects when doing their translation tasks.

Images and sounds can express more information than what the verbal message by itself does, as Pettit (2009: 44) defended. In fact, some information could not be sent to the viewer if there were not images and sounds supporting the transmitted message. That is the reason for the existence of the saying “a picture is worth a thousand words”. For instance, offering somebody more wine with a smile on your face is one thing, saying it with an unfriendly facial expression is another matter. This aspect of meaning cannot be shown just attending to a linguistic or verbal sign. As shown by this example, the image is able to contradict or reinforce what the verbal message is communicating.

Considering Ferklová’s (2014: 7) classification of AVT, the two most popular branches or modalities of AVT are subtitling and dubbing. Nevertheless, apart from these two popular branches, AVT also encompasses some other derivatives which, according to Rica Peromingo, Albarrán, and García (2014: 305), are audio-description,¹ voice-over,² or localisation.³ Dubbing is an AVT mode which involves replacing the SL sound track of the audiovisual text with another track on which dialogues have been recorded in the TL. Therefore, dubbing creates the illusion that the dialogues are originally performed in the TL. The mode of audiovisual translation focused on in this research paper is subtitling. This AVT involves presenting a written text (usually centred in the lower part of the screen trying not to be a nuisance for the viewer) which represents what is being heard through dialogues.⁴ Subtitles also include the many discursive components that create the whole atmosphere of the image shown, such as letters or inscriptions. Moreover, subtitles also express the lyrics of the soundtrack

¹ Audio-description consists in inserting a description or explanation so as to clarify some aspects of what is going on in an audiovisual media for the sake of blind and visually impaired people, owing to the fact that such information cannot be extracted without this additional help.

² Voice-over is a mode of revoicing normally used in documentaries in which there is overlapping of the original soundtrack and another soundtrack in the TL. Therefore, the original version and the translated version can be heard simultaneously, although the volume of the SL track is much lower and the TL track starts a few seconds after the SL one.

³ Localisation consists in an adaptation of any kind of software, videogame, or multimedia file to the country where it is going to be sold, taking in consideration the needs of the market, the linguistic elements, or cultural aspects so as to guarantee the acceptance of the product in that market.

⁴ In Japan, however, film subtitles appear vertically on the right side of the screen.

songs and may offer a description of the information that is extracted through the different noises which can be heard.

All subtitling processes deal with three main elements which are closely related: image, sound and text. These three components must interplay in an efficient way and in synchrony so as to be properly understood. There are many different types of subtitling. First of all, the most well-known type is the transcription of dialogues or any other speech appearing in any audiovisual format into written text. In this specific case, different languages can be involved in the subtitling process. Then, there is a methodology called ‘subtitling by respeaking’, which generates subtitles thanks to a software able to recognise human speech. Consequently, what is being said can be shown in text format immediately. This type of transcription process usually deals with just one language. The third type of subtitles which will be mentioned here is the subtitling for the deaf and hard of hearing (SDH); this translation process can be either ‘intralingual’ or ‘interlingual’, as mentioned in Díaz Cintas and Remael (2014: 15). In other words, the subtitles may reflect what is said in the soundtrack in the same language or it can be translated into a different language. In this kind of subtitling, the viewer can read not only what is being said by the characters, but also an explanation of other aspects related to the audio channel. For instance, this kind of subtitle captures the different noises which can be heard, different accents, the tone of voices, etc. Last but not least, Rica, Albarrán, and García, (2014: 305) refer to another form of AVT regarding subtitles. This is the so-called *fansubbing*, which is basically the “subtitling [process] carried out by non-professionals”. In addition, this non-professional subtitling modality has an equivalent in dubbing, which is called *fandubbing*.

As far as the distribution format of subtitles is concerned, they can be seen in media such as films to be exhibited in cinemas, videogames, television programmes, DVDs, internet, and so on and so forth.

Regarding the many different difficulties that a subtitler may come across throughout the subtitling processes, Marleau (1982: 274) offers a classification where four complications are listed. First of all, it is important to point out the *technical difficulties*. The image is shown on a screen, hence the whole surface of the screen is giving information. Consequently, the space left to the subtitles, usually at the very bottom of the screen, is quite reduced owing to the fact that it interferes with the image. For this reason, each subtitle shown on the screen cannot be longer than two lines; moreover, each of those lines cannot contain more than 39 characters (spaces and

punctuation marks included), and it must last visible for no more than six seconds and no less than one second. In addition, the text has to maintain a temporal synchrony with the image, which means that subtitles have to appear as soon as the character starts speaking and vanish when speech ceases. These temporal and spatial limitations lead to a phenomenon called simplification, which will be discussed later on. Apart from the technical difficulties, the French scholar also mentions the difficulties in relation with *physiological problems*, considering how fast the human eye can read a text. On average, a person can read in one second a word containing from 5 to 8 letters. Nevertheless, during that short period of time, 24 images or frames have been shown on the screen. That is the reason why the length of the subtitles must be adapted to the capacity of reading and, under no circumstances, the subtitle can be on screen for less than just one second. Moreover, the viewer will have to make a considerable effort if s/he is to follow the images, hear the sounds and read the subtitles at the same time. Last but not least, this author listed a last drawback of subtitles, the *artistic* or *aesthetic problem*; considering that subtitles damage the quality of the image and that speech has to be simplified sometimes. For both reasons, the French scholar believes that both the quality of image and speech is affected by subtitles.

3. Linguistic corpora

A corpus may be defined in linguistics as a systematic collection of empiric data of language in use. It can consist of spoken language, written texts, or both. In addition, corpora can be focused on one specific genre, such as science or literature, or they can include general language, extracted from whatever field of study or daily language. The process of collecting the data nowadays is computerized, and the samples are stored virtually. This kind of corpora are compiled thanks to the labour of computers with the help of specialised software that offers the advantage of aligning and ordering the samples of language as well as searching for specific words or sentences in the collected data.

These compilations of linguistic data can be extremely useful in many different fields of study; owing to the fact that it is a researching tool containing a huge amount of information. How this tool works is quite simple: it offers the user a search bar where the word or words that are to be studied can be introduced. Then, the programme searches it in the collected data and shows the result, which basically consists of the different contexts in which these words appear. For this reason, as Perez-Avila (2006) establishes in his work, it is a tool of interest in teaching, as linguistic corpora show the different collocations and contexts of words or complex structures. Consequently, teachers can use them to show how some specific aspects of language are used, as they only have to search for the word or structure and they will find thousands of examples of how this form is used, the elements that could precede and follow it, and so on and so forth. As Biber, Conrad, and Reppen, (1998: 1) claimed, linguistic corpora allow the researcher to study the “actual language used in naturally occurring texts”. That is the reason why different studies considering language use and interested in investigating one particular linguistic structure use linguistic corpora as a researcher tool. As a case in point, Rica Peromingo (2014) studied how discursive markers in English subtitles are translated into Spanish using a specific linguistic corpus.

Following Nesselhauf's (2005: 2) classification, there are some different types of linguistic corpora regarding the number of languages that the corpus deals with, the nature of the compiled information and how it is shown, as well as the specific advantages that each corpus offers. With respect to the type of language, linguistic corpora can be separated into three different subtypes. First of all, there are linguistic corpora that are aimed at studying spoken language; these are the so called ‘spoken

corpora'. *The Santa Barbara Corpus of Spoken American English*⁵ is an example of this type of corpus; it is a small downloadable corpus, formed by approximately 250 thousand words extracted from audio files which are at the user's disposal together with the transcription of the audio and the correlative timestamp of each fragment. Some other corpora are devoted to analyse written texts only. These are called 'written corpora' and can be illustrated, for instance, by *Frantext*⁶, composed by almost 150 million words belonging to 500 French literature works written between the 18th and the 20th Century. Finally, the vast majority of linguistic corpora include both written and spoken samples of language as *The Corpus of Contemporary American English*⁷ (COCA) does, which encompasses samples of language in use extracted from many different sources, no matter whether the language is spoken or written. The latest corpus offers the ludicrous number of 520 million words. It is by far the widest English linguistic corpus publicly available. The creator of this electronic platform was Mark Davies.

As regards the amount of languages concerned, there exist two different subtypes of linguistic corpora: monolingual and multilingual. Whereas 'monolingual corpora' include extracts belonging to a single language, 'multilingual corpora' deal with more than one language. Within the monolingual category, two different subtypes of monolingual corpora can be distinguished. On the one hand, there is a kind of corpus whose main aim is to represent all the structures and aspects of one specific language. This type of corpus is generally called 'reference corpus' and is normally used as *tertium comparationis* in translation studies. One of the most outstanding reference corpora is the *British National Corpus*⁸ (BNC), which offers an updated collection of 100 million words of language in use, both spoken and written language from a diverse variety of sources. Reference corpora deal with language in general, including different genres and registers. On the other hand, 'specialized corpora' include language samples from a specific subject area. The *Vercial Corpus*⁹ exemplifies this kind of corpus. It provides more than 14 million words related to Portuguese literature between the 16th and the 20th century. This corpus leads directly to the next category of monolingual corpora, as the *Vercial Corpus* does not belong only to the subtype of specialized

⁵ <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

⁶ <http://www.cnrtl.fr/corpus/>

⁷ <http://corpus.byu.edu/coca/>

⁸ <http://www.natcorp.ox.ac.uk/>

⁹ <http://www.linguateca.pt/acesso/corpus.php?corpus=VERCIAL>

corpus, but also to the ‘historical corpus’ subtype, as it deals with language in history. These tools are useful for diachronic and synchronic studies, as they enable to study how language has developed through the passing of time. Furthermore, any stage of language can be studied in isolation as well by means of this kind of linguistic corpus. In addition, there are also ‘corpora of the present day language’, which focus their attention only on current language use, such as the previously mentioned *Corpus of Contemporary American English* does. The next type of corpus that will be referred to in this paper is the ‘translational corpus.’ This kind of tool allows the possibility of researching and analysing translated fragments of spoken or written language. One of the most prominent corpus of this type is the *Translational English Corpus*¹⁰ (TEC). It contains around 10 million words from written texts translated into English from a variety of languages.

Leaving aside the many corpora that handle only one specific language, another type of linguistic corpus much closer to the interests of this paper is the so-called multilingual corpus. As their name indicates, multilingual corpora, unlike monolingual corpora, deal with more than one language. Within multilingual corpora, the specific type of corpus which contains samples from two different languages is the so-called bilingual corpus. A clear example of bilingual corpus is ACTRES,¹¹ which offers more than 4 million words from original English texts and their translations into Spanish and vice versa.

Owing to the fact that a research in subtitling translation from English into Spanish will be shown after this theoretical approach, the kind of corpus needed is a bilingual one which centres its attention on the translation of subtitles. Taking this into account, there is a specific type of bilingual linguistic corpus which is very useful for translation studies: ‘the parallel corpora’. This kind of tool offers both the text of the SL and its TT counterpart in an aligned way. Therefore, it allows studying the different translation strategies used in the translation process in a deep way. Due to the current increase of translated documents available on the Internet, as mentioned in Tiedemann (2007: 1), these corpora handling multilingual material are getting remarkably popular because of their usefulness for translating purposes. In fact, this essay is aimed at studying translation processes in film subtitles. Consequently, some parallel corpora which deal with this mode of AVT have been taken into consideration.

¹⁰ <http://www.alc.manchester.ac.uk/subjects/tis/research/projects/tec/>

¹¹ <http://actres.unileon.es/inicio.php?elementoID=12>

As mentioned above, all the process of compiling the data for linguistic corpora is done electronically thanks to specialized software. Xiao and Wang (2009) explain the process of constructing an aligned parallel corpus focused on film subtitles. As highlighted by them, during this compiling and organising process, human beings do not perform any particular task but programming the software and supervising that everything is working properly. Some online tools offer the chance of compiling a corpus consisting of STs and their TT counterparts. As an example, the project SUMAT¹² is an online service that offers this service. It is aimed at increasing the efficiency and accuracy of subtitling processes thanks to machine translation technology. In order to ensure that the subtitle data is of high-quality, SUMAT works in collaboration with professional translation companies, such as *Deluxe Digital Studios*¹³ and *Titebild Subtitling and Translation*¹⁴, formed both of them by professional translators capable of handling translation software. The contribution of this tool, together with the help of the previously mentioned companies, has compiled more than 1 million professionally translated subtitles in more than 50 different languages which are organized for researching purposes.

To start with, there is a parallel linguistic corpus dealing with subtitling that deserves some attention owing to the many advantages and possibilities that it offers. It is the *Veiga Corpus*¹⁵. In Sotelo (2012), the designer of this corpus explains its features as well as the creation process. The *Veiga Corpus* is a multimedia linguistic corpus that offers the subtitles of 24 English films, both in English and Galician. The user can search any word or sequence in both languages, and the contexts where the word or sequence of interest is uttered will be shown (See Figure 1 below). Nevertheless, it is important to bear in mind that this platform is still under construction. That is the reason why the amount of samples is quite limited if compared to other more developed corpora, as this one has approximately 300 thousand words. The exclusiveness of this platform resides mainly in the feature that it gives the researcher the advantage of watching the footage where the word or expression that has been searched is uttered. (See Figure 2 below). For this reason, apart from helping researchers in their translation studies, it also provides practical evidence.

¹² <http://www.sumat-project.eu/>

¹³ <http://www.bydeluxe.com/>

¹⁴ <http://www.titelbild.de/>

¹⁵ http://sli.uvigo.es/CLUVI/vmm_en.html

Look, why the fuck are you gonna bust ¶ my balls about this now?	¿Por que me queres foder ¶ agora con isto?
Now, it's so fucking cold there, ¶ they can only pull the diamonds...	Vai tanto frío que só sacan ¶ diamantes
Tell me this ain't fuckin' ¶ doable.	¶ Dime que non é posible.
Well, fuck him.	Pois que o fodañ.
Where the fuck you been, man?	-De onde vés?

Figure 1. Search of the word “fuck” in the *Veiga Corpus*.



Figure 2. Example of the English word “fuck” in context and how it has been translated into Galician extracted from the *Veiga Corpus*.

However, the target language of interest in this research is not Galician but Spanish. Next and for this reason, the linguistic corpora that are going to be dealt with below are devoted to translating English subtitles into Spanish. For instance, CORSUBIL (*Corpus de Subtítulos Bilingües inglés-español*) is a parallel corpus that encompasses approximately 18 million words extracted from film subtitles. It is analysed with a wealth of detail by Rica Peromingo, Albarrán, and García (2014: 312). This corpus is rather unusual because it is formed by two subcorpora, each of them compiling around 9 million words: CORSUBILES (*Corpus de Subtítulos Bilingües en Español*) and CORSUBILIN (*Corpus de Subtítulos Bilingües en Inglés*). The latter one contains original English subtitles and the former one the translation of those ST subtitles into Spanish. The combination of both form the CORSUBIL corpus, which is nowadays growing and adding more data to the platform. Nevertheless, it presents some drawbacks. The data are not offered in a parallel way; for this reason, if the SL and its translation are to be compared, all the compiled data of both subcorpora have to be managed. Moreover, the public access to this corpus is restricted and it is not available online. The privacy of this corpus is not a unique feature. In fact, many small subtitling

corpora collected by certain researches are not available for public consultation. For instance, there is a corpus composed by extracts from subtitles of the TV sitcom *Friends*, both in English and Spanish. The researcher and the compiler of this corpus is Romero Fresco (2005), who studied the translation of phraseological idiomaticity.

The parallel corpus which has been selected to accomplish the purpose of this research, due to the fact that it has been demonstrated to be the most efficient and accurate one for achieving the goal of this paper, is the *Open Parallel Corpus* (OPUS)¹⁶, described by Tiedemann (2009). This corpus is an expanding collection of original English subtitles and their translations into several different languages. This linguistic corpus is probably the largest collection of subtitles, among the free parallel linguistic corpora. As stated in Tiedemann (2010: 2214), “[t]he overall goal of the OPUS project is to make parallel resources freely available.” For this reason, all the data are publicly available and can be consulted and downloaded for free.¹⁷

OPUS encompasses texts from a wide amount of fields of study from 90 different languages. Within the different domains that the compiled data deal with, a special mention should be made to administrative and legislative texts and also to movie subtitles. Among others, the major parallel corpora included in this wide project are: ECB¹⁸, which encompasses about 30 million sentences in 19 languages from the *European Central Bank*; MBS¹⁹, which is a compilation of French and Dutch newspapers, containing about 10 million words; and all the updates of the *OpenSubtitles* corpus²⁰, which is the subcorpus used in this study. It consists of 22 million sentence alignments from more than 1.3 million subtitles from films in more than 54 different languages. The version which has been used for this aim has been the updated 2011 version. This version counts on the help of a search interface called CQP²¹. This programme allows the researcher to look for any word or sentence within the SL corpus, as well as to see how it has been translated, providing both the original text and its translation in a parallel format. A further analysis of this subcorpus is shown in the methodology section along with how it has contributed to the development of this project.

¹⁶ <http://opus.lingfil.uu.se/>

¹⁷ For instance, the updated *OpenSubtitles* corpus is freely downloadable from this webpage: <http://opus.lingfil.uu.se/OpenSubtitles2016.php>

¹⁸ <http://opus.lingfil.uu.se/ECB.php>

¹⁹ <http://opus.lingfil.uu.se/MBS.php>

²⁰ <http://www.OpenSubtitles.org/en/search>

²¹ <http://opus.lingfil.uu.se/cwb/OpenSubtitles/frames-cqp.html>

4. Pragmatic functions and translation of taboo words

Since, in this study, we are considering the translation of language in use, the only appropriate model would be that which considers not only its linguistic form but also the context and the communicative intention. Consequently, we will take pragmatic factors into consideration when analysing the translation of taboo words.

4.1. Pragmatic functions of taboo words

This pragmatic approach has to be considered when translating what Ferklová (2014: 17) refers to as ‘bad words’, as swearing is influenced by sociocultural and contextual aspects as defended by Jay (2008: 272; 2009: 154) more than once. This category is formed by vulgarisms, taboo words, profanities, dirty words, slang words, and swear words²², all of which are intimately linked with culture. Hence, the intention and consequences of using them are normally different when two different cultures are taken into account. That is the reason why, as Jay (2008: 269) mentions, the phenomena of impoliteness, swearing and rudeness “are impossible to define universally because all are culturally and personally determined”. When translating these kinds of words, rarely is it seen that a specific structure is translated identically in each and every context where it appears. This happens because each language has its specific ways to express different meanings and intentions.

First of all, it is essential to distinguish between the concepts of taboo, slang, and swear words. To start with, the very last concept is similar to offensive words. These words are used, as a general rule, to scorn someone. Then taboo words are not just insults but proscribed words which are not accepted in a given society. The fact of uttering this kind of word may have as a consequence a great repulsion and an offense to the listener. Finally, slang words and expressions correspond to informal language. They are used more frequently in oral speech rather than in written texts. Nevertheless, it is not unusual to find them in written language. Each culture has its own preferences regarding these kinds of words and expressions. What is established as a common expression in one language can be seen as an affront in another one. For instance, the Spanish word *coño* is commonly used as interjection expressing astonishment; however, it is unsuitable to utter the English equivalent of this word in this context. In fact,

²² Throughout this work I have used the concepts ‘swear word’ and ‘taboo word’ to allude to potentially offensive language.

English speakers normally avoid using this word when in public. Likewise, words which are considered as taboo or swearwords in the source culture may be translated as words that are not taboo in the target culture.

But, why do speakers use taboo words and swearwords? To explain the several reasons for using them, the different taboo categories established by Ferklová (2014: 17) are going to be presented. As mentioned above, each culture has its own taboo matters. Nevertheless, as a general rule, the majority of languages share certain categories. First of all, there is a large group of taboo words which is connected with sex and every aspect regarding it, such as sexual organs, masturbation, homosexuality, and so on and so forth. Another large category, connected with the human body as well, encompasses words and expressions denoting bodily waste and fluids. As these substances generate repulsion, they are commonly unacceptable. In some languages, some words in relation with food have become taboo; mainly because of health protection, as some animals could carry some illnesses, and because of religion as well, as eating some kinds of food is not permitted by some certain religions. Mentioning somebody's skin-color could hurt feelings, and for this reason, it can also be said that ethnicity and race are two other sources of taboo words. In addition, many aspects related to religion should not be mentioned in certain situations. For instance, categorizing someone as being a member of a religious group can be considered insulting. Moreover, in religions such as Christianity, uttering the name of God 'in vain' is seen as sinful. Two subgroups of taboo words concerning religion can be distinguished. On the one hand, celestial words, which are those that make reference to good things, such as 'heaven', or 'Jesus Christ'. On the other hand, there are diabolic words, such as 'hell' or 'demon'. Leaving aside religion, there are some other offensive words dealing with disabilities, both mental ('retard') and physical ('lame'). The last but certainly not least semantic field that encompasses a wide amount of taboo words in several cultures includes illnesses and death. This fact is based on the human's fear to die as well as superstition concerning the spiritual world.

This previous classification established by Ferklová (2014: 17) reveals the different semantic fields to which the denotative meaning of taboo words usually belong. Nevertheless, swearwords are commonly uttered to state connotative meaning, such as how we feel or how we are. The connotative meaning that is intended to be uttered depends mainly on the intentions and the attitude of the speaker. What is more, if the speaker wishes to offend someone, almost every word could turn into offensive, in

spite of having a different denotation. This aspect is clearly linked with the pragmatic of those words. The pragmatic functions that swearwords can perform are quite varied. These are used, according to Wang (2013: 72) and Pluszczyk (2015: 110), to express negative emotions, emphasis, expressing humour, showing social relationships, and for aggressive purposes. In brief, as Jay (2009: 155) asserts, the different “reasons for using [...] taboo words depend on the conversational goals of the speaker”.

The most important pragmatic factors that determine the consequence of uttering a swearword are, as Jay (2008: 273; 2009: 154) illustrates, the intention of the speaker together with the relationship between the characters involved in the speech act within the social location where the dialogue takes place. The different feelings that the speaker is showing with the speech is a crucial factor to take into account when translating the message. As stated by Pujol (2006: 123), Wang (2013: 72) and Pluszczyk (2015: 110), a wide amount of emotions can be expressed through taboo words. Pujol (2006:123) defends that taboo words cannot be classified in only one pragmatic category according to the emotion which the speaker tries to show. On the contrary, he justifies that they can overlap and express more than one feeling.

4.1.1. Pragmatic functions of ‘fuck’ and its derivative ‘fucking’

First of all, owing to the outstanding importance of the word ‘fuck’ in the English language considering its frequency of use, the pragmatic functions of this word are going to be illustrated first. This taboo word does not always imply its literal meaning, “to have sex”. So many different connotative meanings apart from the denotative one are expressed with this word and its variants. The main communicative functions of ‘fuck’ are illustrated in the table below:

Uses of ‘fuck’	
Denotative meaning	<i>Excuse me, do you fuck?</i> <i>Even when I fuck I talk.</i>
To express emotion	<i>Stupid fucking American people.</i> <i>What the fuck happened to my car?</i>
Emphasis	<i>I’ve fucking won.</i> <i>I’m fucking great!</i>
Offending	<i>Come on in, you fuckers.</i> <i>You dumb fuck!</i>
Solidarity	<i>Can’t believe you guys are turning me into a fucking boy scout.</i> <i>This guy is the fucking master, don’t you see?</i>

Table 1. General uses of the taboo word ‘fuck’ and its derived words.

This taboo word can be used in many different contexts and with many different purposes. First of all, one of its main functions is to express emotion. It is generally uttered for expressing negative and strong emotions, such as anger, anxiety, fear, surprise, frustration, annoyance, or fury. Apart from this use, the f-word, and mainly its -ing form, can be uttered to emphasise the element it goes with. Thus, this taboo word mainly premodifies nouns (“Stop the fucking car.”), adjectives (“I’m fucking blind.”), verbs (“Everyone you’ve ever known, everyone you’ve ever fucking loved.”), and whole sentences (“what the fuck are you assholes doing?”). When used as intensifier, this word barely implies a semantic connotation. In addition, obviously, this swearword can be used to insult someone. Hence, it is certain that it may imply a negative meaning. To conclude with the categorization of uses of ‘fuck’, it is essential to discuss the ability of this word to express solidarity and identity. In this sense, when this taboo is uttered with this purpose, it implies that the person to which the message is addressed belongs to the same social group and possesses an intimate membership.

4.1.2. Pragmatic functions of ‘shit’

As mentioned above, ‘shit’ is the second most frequent four-letter word used in the *OpenSubtitles*. Due to the standardization of this word, it can be seen in G rated films, as it is kind of impolite but at least not considered to imply profanity or to have strong connotations. This word, and also its morphological variants, is nowadays commonly used in different collocations. In addition, this concept, together with all its morphological variants, can be used with many different purposes. As the sometimes offensive word ‘fuck’, the taboo word ‘shit’ can also be used figuratively expressing different meanings apart from the literal one. These connotative meanings are listed below:

Uses of ‘shit’	
Denotative/literal meaning	<i>Well, I guess shit doesn’t float after all. That dragon smells like shit.</i>
To express emotions	<i>Don’t move, shit! Oh, shit!</i>
Offending	<i>Stop the car, you piece of shit. Give me that, you shithead.</i>
Informal way to refer to any object	<i>You don’t believe that shit for a minute. I’ve used that shit.</i>

Table 2. General uses of the taboo ‘shit.’

As regards the connotative uses of this word, three different branches can be discussed. To start with, it is common to express emotions using this word, usually negative ones as happened with the word ‘fuck.’ In this case, the speaker is generally carried along by his/her emotions and utters the swearword as an emotional response. Still, even though it is pronounced automatically, it is normally considered as impolite as well. It is also common to use this word and its morphological variants in order to insult others. For this purpose, the uttered word is commonly a variant of the base form, such as ‘shitface’, ‘shitty’, and ‘shithead’. This swearword can also be included in a set phrase for implying this intention, as the example “stop the car, you piece of shit” illustrates. Nevertheless, this word is not commonly used if the speaker wants to give emphasis to certain elements, as the word ‘fuck’ frequently does. The last utility of this word is to make reference to any object in a kind of pejorative way. In this specific case, it can be said that it functions as a pejorative and informal synonym of the words ‘thing’ or ‘stuff’.

The way in which both words can be used is quite different. Nevertheless, even though both words are labelled as taboo, they do not always express a damaging or pejorative meaning. As both ‘shit’ and ‘fuck’ are content words, they have a semantic definition; even though their denotative meaning is the least common in film dialogues. Thus, only a quite limited percentage of these words imply, in the case of the verb ‘fuck’ and its variants, the action of having sex; and in the case of the noun ‘shit’ and its variants, the concept of bodily waste. In the case of the word ‘fuck’, it can be used with a figurative meaning for instance in phrasal verbs like ‘fuck over’ or ‘fuck around’, which can be illustrated by the following examples: “Don’t fuck around with the boss’s daughter” and “I don’t wanna have to fuck you up, Smokey.”

4.2. Translation of taboo words in subtitling

It is common to find that there is no equivalent for certain swearwords between two particular languages, as Surià (2012) highlights. According to this scholar, this has to do with the fact that taboo words are commonly culture-specific. For this reason, if there is no equivalent and a literal translation is carried out, the result may sound quite unnatural to the TL audience. Furthermore, when translating subtitles, professionals have to take into account that changing the mode, from speaking to written language, may cause that swearwords could be perceived as more offensive and rude. In order to prevent this

from happening, subtitlers have to evaluate the possible impact that the translation may cause on the target-culture audience beforehand. Afterwards, when this has been done, the translator can continue with the process, rendering the swearword by means of an equivalent that is sure to be appropriate for this specific culture and the context where it is used.

As previously mentioned, the transfer of swearwords between two languages will depend on the social environment where the character is uttering the given expression. It will also be influenced by the situation that brought this person to express him/herself in that way, as the meaning of the swearword will be determined by the situation. For instance, saying the word ‘shit’ because of surprise, such as in “Holy shit!” should not hurt someone’s feelings; on the contrary, saying it out of anger to address somebody could offend the hearer, as in “You fucking piece of shit”. Last but not least, the third aspect which may condition the transfer is the addressee, to whom the message is addressed. It can be concluded that taboo words are culture-specific. For this reason, as mentioned in Díaz Cintas and Remael (2014: 11), “[t]he one-to-one translation approach loses all validity in our field”. Translating literally without taking into account some cultural and pragmatic factors is rarely a valid and accurate technique, especially if dealing with swearing. In this case, translators have to follow some techniques to translate this lexis accurately. The five translation strategies of taboo words that are going to be discussed in this study are: literal translation, omission, softening (or de-swearing), change of swearword, and transposition. The first three strategies have been discussed and defined by Han and Wang (2014: 9). The main strategy used by subtitlers is the literal translation; which implies using the formal equivalent of the swearword in the TL. For instance, a clear example of literal translation is “He usado esa mierda”; which has been literally translated from “I’ve used that shit”. Secondly, the ‘omission’ strategy implies the deletion of the word in the TT. In this sense, taboo is not present in the translated text. On some occasions it may be considered as a sort of censoring, as some information is not shown intentionally. For instance, if the slang or swearword to be translated performs the role of filler in communication it is often omitted. This is because these words are rarely fundamental for the understanding of the plot. This fact can be seen in the following example: “Are you fucking mad?”, which has been rendered into Spanish as “¿Estás loco?”, ignoring the presence of the swearword ‘fucking’, as it is not critical for extracting the meaning of the sentence. Then, there is a strategy that implies translating the swearword into a

more neutral or softer word. That is to say, the word is not omitted but softened. This strategy is generally called softening. However, some experts call it de-swearing or mollification as well. So as to clarify this concept, the following example of how the word ‘fucking’ has been translated can be considered: in the original subtitle, we find “You’ll never be the fucking man”, which has been translated into Spanish as “Jamás estarás a la altura”. The meaning is the same, but in the Spanish translation it has been expressed in a less aggressive way than in the original subtitle. Afterwards, there is a strategy that is represented in the samples of *OpenSubtitles* which we have called ‘change of swearword’. It consists basically in rendering the taboo word by means of a different swearword, preserving its taboo component. Finally, Han and Wang (2014: 9) mentioned another common strategy in the process of translation of subtitles, called ‘category shift’ or, as other scholars call it, ‘transposition’. This is not so frequent when translating English to Spanish or vice versa, as when working with these two languages the word-class is normally preserved. Nevertheless, although it is minimally present in the studied translated subtitles, it deserves to be mentioned and exemplified. It consists of changing the grammatical category of the SL word or sentence. In these cases, the meaning is not softened or omitted, but remains the same. Below, in the result section, these strategies will be exemplified and discussed using samples of subtitles from *Opensubtitles*.

5. Methodology

After examining the characteristics of the English taboo words, it is time to illustrate how the translation of these elements is done according to the samples of data available in *OpenSubtitles*. The first step to prove this is to present the methodology followed to carry out this study. First of all, the general and specific aims of this research will be discussed. Afterwards, the tool used for extracting the data of interest will be described. Thirdly, the five stages followed to achieve the goals of this research are presented.

5.1. Objectives

General objective

- To accomplish a descriptive study on the translation of taboo words in film subtitles, using the *OpenSubtitles* corpus as research instrument.

Specific objectives

- To identify the different word-classes fulfilled by the taboo words ‘fuck’, ‘shit’ and their morphological variants.
- To analyse how if the variable grammatical class determines the choice of translation strategy to translate the taboo words focused on.

5.2. Research instrument

The linguistic corpus used for this purpose is, as mentioned above, a subcorpus from OPUS that encompasses film subtitles. Tiedemann (2007; 2009; 2010) highlighted the usefulness of this programme. The present paper aims to analyse sequences of English subtitles containing taboo words and their Spanish translations. For this purpose, the swearwords ‘fuck’ and ‘shit’ together with their morphological variants are going to be closely examined; trying to extract some generalizations about how they are rendered in Spanish subtitles. As stated in the introductory part of this study, each and every example has been taken from the *OpenSubtitles* corpus thanks to a search interface provided by the OPUS parallel corpus.

The immense majority of the subtitles included in this corpus belong to the genre of spoken language as subtitles are normally written transcriptions of film dialogues. That is the reason why we find many incomplete sentences, self-corrections, false starts, and many other characteristics of spoken language. However, the kind of language

represented in the original subtitles also differs a little from spoken language, as film screenplays are written by professional script writers.

The screenshot shows the CQP interface. At the top, there is a search query input field containing "viol.*". To the right of the input field is a "Sort =" dropdown menu set to "unsorted", and a "Reset Form" button. Below the input field are three buttons: "Run Query", "Distribution", and "Frequencies". Below these buttons is a "Display:" section with a "context =" dropdown menu set to "sentence". At the bottom, there is an "Alignments:" section with a row of checkboxes for various languages: bg, cs, da, de, el, es, et, fi, fr, he, hr, hu, is, it, ja, lt, nl, no, pl, pt, pt_br, ro, ru, sk, sl, sv, tr, zh.

CQPDemo: Search the OpenSubtitle corpus with CQP query.

Will display short summary and some help here.

Figure 3. Welcome interface of CQP for searching into linguistic data of *OpenSubtitles*.

The CQP tool manages the data from *OpenSubtitles*, allowing the researcher to check how film subtitles have been translated. Each subtitle consists originally of one or two lines, which is the maximum length that should be projected on the screen according to the physiological, psychological, aesthetic, and technical considerations previously discussed. Nevertheless, CQP can modify the length of both the subtitle in the SL and its translation so as to suit user's preferences only by a click in the gap situated at the right side of "context", as illustrated in Figure 4 below. In addition, by clicking in the hyperlink "context" placed at the left side of the subtitle, a pop-up window will appear showing the context of the SL subtitle. There, the researcher is shown the whole context where the word or clause is used, as can be seen in Figure 5.

The screenshot shows a close-up of the "Display:" section. The "context =" dropdown menu is open, showing a list of options: "sentence", "5 words", "sentence" (highlighted in blue), "2 sentences", "4 sentences", "1 timeslot", and "2 timeslots". To the left of the dropdown menu is the "Alignments:" section with a checkbox for "bg".

Figure 4. How to select the context where the searched word is wanted to be shown.

The screenshot shows a subtitle entry: "10. an_american_in_canada, movie 3443, sub 211529". Below the entry is the subtitle text: "context So I'm off your shit list ?". The word "shit" is highlighted in yellow. Below the subtitle text is a large block of text showing the context of the subtitle. The context text is: "Now , I may not be a mind reader .but I think I know what you were gonna say .You want me to be team captain . I don't think I've been here long enough to have gained the respect and admiration of the other guys . But , um . What the hell . I accept .
You know , that's just a crazy enough idea to actually work . - Captain Wagstaff . - Really ? So I'm off your shit list ? No , you puke . I've just spent more than a day and a half trying to mold you into an Olympic- caliber snowboarder . I just realized that I've been polishing a turd .
You are fuckin' expelled . Huh . Well , thank you for you support , sir .Who would I see about getting a tuition refund ?You're lookin' at him , scumbucket .Guess what ? I'm going to impound that homo- blowjob slush fund .All right . I'm leaving ."

Figure 5. Word searched for together with its context

Moreover, as it is a parallel corpus, it offers the option of checking how the subtitle has been translated into another language. The two following examples in

Figure 6 show two subtitle samples together with their aligned translation extracted from the English-Spanish corpus of *OpenSubtitles*:

102. swordfish, movie 1756, sub 125123	es
context They make shit .	Hacen pura mierda .
103. swordfish, movie 1756, sub 125123	es
context Unbelievable , unremarkable shit .	Mierda increíble y mediocre .

Figure 6. Two examples of a search in *OpenSubtitles* using CQP.

5.3. Research stages

This study aims to provide a descriptive research into the different strategies adopted when translating taboo words from English into Spanish in film subtitles. To achieve this goal, some stages have been followed by the researcher, which can be summarized as follows:

1. Searching of the taboo words in *OpenSubtitles* considering English as SL and Spanish as TL.
2. Classifying these taboo words together with the aligned translation considering its grammatical category.
3. Analysing the different translation strategies carried out in the corpus according to Han and Wang's (2014: 9) classification.
4. Conveying all the data of my classifications of the taboo words and their translations into Spanish to an Excel spreadsheet for further analyses.
5. Extracting and interpreting the results of the different translation techniques for translating swearwords as well as the possible effect of certain variables on the choice of strategy.

This investigation is going to focus its attention only on the words 'shit', 'fuck', and its morphological variant 'fucking' owing to their significance in the English language. The translation strategies at word-level used in the corpus to render these words have been classified in an Excel spreadsheet, which allows to extract the frequency for each strategy as well as the effect of variables such as the grammatical category. The table below is extracted from the Excel spreadsheet where all the swearwords have been compiled and analysed.

1	SL Taboo	Gram. category	TL translation	Translation strategy
2	fuck	interjection	mierda	change of swearword
3	fuck	interjection	-	omission
4	fuck	verb	joder	literal translation
5	fuck	verb	jodes	literal translation
6	fuck	interjection	-	omission
7	fuck	interjection	-	omission
8	fuck	adjective	capullo	change of swearword
9	fuck	interjection	-	omission
10	fuck	verb	follen	literal translation
11	fuck	adjective	tonto	softening
12	fuck	interjection	-	omission
13	fuck	verb	jode	literal translation
14	fuck	adjective	gilipollas	change of swearword
15	fuck	interjection	-	omission
16	fuck	interjection	-	omission
17	fuck	verb	jodas	literal translation
18	fuck	verb	juegues	softening
19	fuck	verb	joderme	literal translation
20	fuck	interjection	joder	literal translation

Figure 7. Examples extracted from the Excel spreadsheet designed for the analysis of the data.

6. Results

As previously stated, this BA dissertation is aimed at giving a descriptive analysis of how the word ‘shit’, ‘fuck’, and its morphological derivative ‘fucking’ is rendered from English to Spanish subtitles. That is the reason why this section provides an analysis explaining the different translation techniques employed according to the different factors that may influence the rendition. In this section, the different strategies used for translating subtitles in the corpus analysed in this study are discussed and illustrated with examples. The different factors that may have an effect when translating taboo words have been considered as well. With this purpose, examples from the *OpenSubtitles* corpus are going to be examined.

Examining the data, it can be assured that the most frequent taboo word present in the corpus is, by far, ‘fuck’ together with its several morphological variants. This word appears 3,776 times in the *OpenSubtitles* corpus. The second word in the rank of the most used swearwords is ‘shit’ and its variants, with a total of 2,039 matches in the same parallel corpus. The f-word is particularly common despite of its taboo content in oral speech, as declared by Pujol (2006: 122). This does not mean that it is never included in written texts, but it is quite less frequent than in spoken language. Even though this word is used as an expression in our daily lives without trying to damage anyone, it is supposed to be censored on TV, although it is not difficult to hear it or its morphological variants every now and then. On films, Parental Guidance Suggested (PG rated) films included, it is quite common to find this word. It should not appear in General Audiences films (G rated), as all ages are allowed to watch them. In Wikipedia, there is an interesting section entitled “List of films that most frequently use the word ‘fuck’”²³, where a ranking containing several films where this swearword appears is shown. In fact, the subtitles of some films that are included in this rank are present in the corpus. For instance, the subtitles of the film *Monster* and *Snatch*, which are at the 80th and the 109th position respectively, can be checked if wished. It is unquestionable that this word is the most frequent swearword in English films, followed by the taboo ‘shit’.

²³ https://en.wikipedia.org/wiki/List_of_films_that_most_frequently_use_the_word_%22fuck%22

6.1. General strategies for translating taboo words

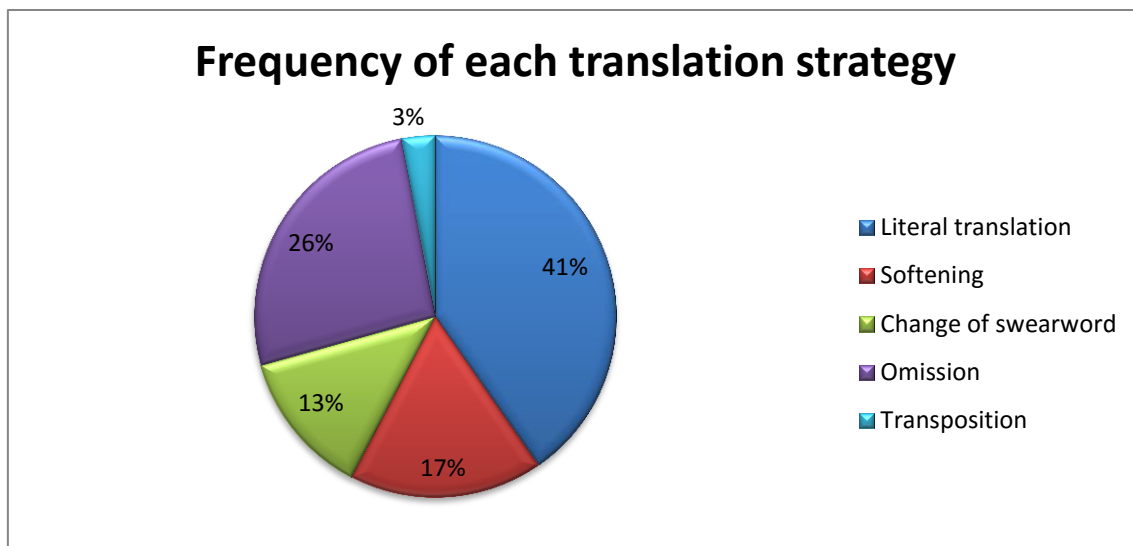
After analysing the entire amount of data offered by *OpenSubtitles* regarding the results obtained by searching ‘fuck,’ its morphological variant ‘fucking,’ and ‘shit’ in the corpus, it has been demonstrated that the previously defined translation strategies have been used to render the above-mentioned taboo words in the subtitles compiled in the corpus.

The five translation techniques for translating taboo words present in the corpus, ranked according to their frequency, are: literal translation, omission, softening (or de-swearing), change of swearword, and transposition. Below, a general overview of these translation strategies found in the studied corpus is offered together with illustrative examples extracted from the corpus:

Translation strategies	Subtitles where the strategy have been applied
Literal translation	SL: <i>You did that on purpose just to fuck me over.</i> TL: <i>Lo hiciste a propósito sólo para joderme.</i>
Omission	SL: <i>Tell him to get his fucking ass back here!</i> TL: <i>¡Dile que venga!</i>
Softening	SL: <i>You don't believe that shit for a minute!</i> TL: <i>¡No creerás en esas tonterías!</i>
Change of swearword	SL: <i>What the fuck?</i> TL: <i>¡Que te jodan!</i>
Transposition	SL: <i>I couldn't give her a good fuck</i> TL: <i>No pude follarmela</i>

Table 3. Translation strategies found in the corpus with a representative example.

The strategy used more frequently by subtitlers according to the results extracted from this analysis is ‘literal translation,’ reaching 41% of the cases. The second strategy most used within this corpus, with 26%, is ‘omission’, followed by ‘softening’, which represents 17%.

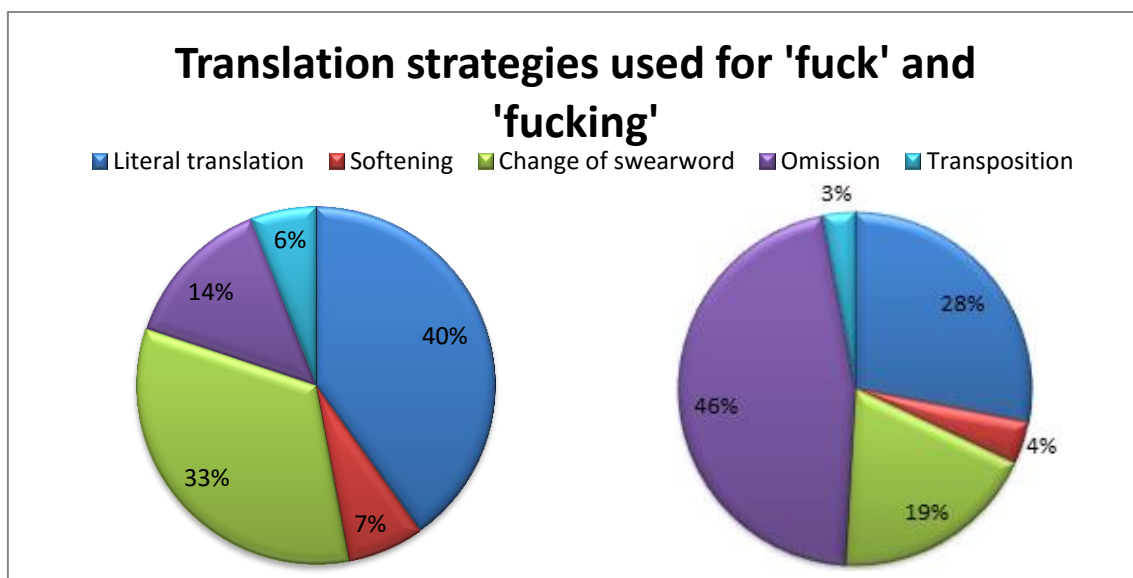


Graph 1. Translation strategies of taboo words identified within *OpenSubtitles*.

The chart above illustrates the percentages of use of these translation techniques. Analysing the data in more depth, the two subsections below focus on the translation strategies used in the rendition of each of the studied taboo words.

6.1.1.1. General translation strategies used for translating ‘fuck’ and ‘fucking’

The graph and the table below illustrate the frequency of the translation strategies used for translating either ‘fuck’ or its morphological derivative ‘fucking.’



Graph 2. Translation strategies used for translating ‘fuck’ (left side) and ‘fucking’ (right side).

	Literal translation		Softening		Change of swearword		Omission		Trans-position		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
fuck	532	40.0	93	6.9	439	33.0	186	13.9	79	5.9	1,329	100
fucking	576	27.9	82	3.9	391	18.9	947	46.0	62	3.0	2,058	100

Table 4. General translation strategies used to translate ‘fuck’ and ‘fucking.’

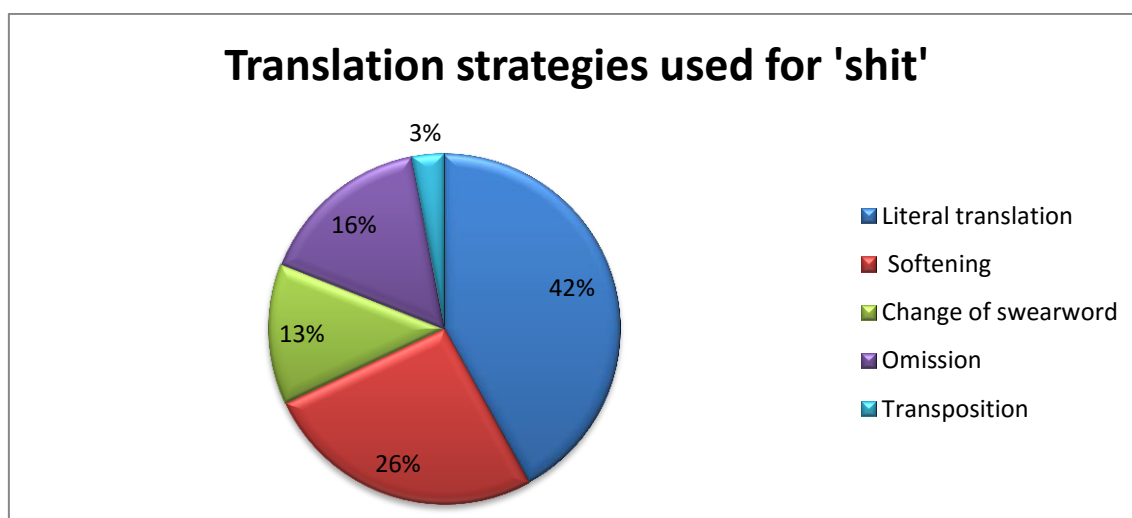
According to the results of obtained by the present study, it can be determined that the translation strategies are used in a different way to render ‘fuck’ and its derivative ‘fucking’. First of all, the base form of this taboo word is translated literally as in (1) in 40.0% of its 1,329 occurrences. Afterwards, example (2) shows a case in which ‘fuck’ is softened as less pejorative word. This accounts for 6.9% of the times that this word appears in the corpus. Then, in 33.0% of the cases, the subtitler decided to change the swearword ‘fuck’. In these cases, a different taboo word but of the same grammatical category is used in the TT as in (3), where the translator opted to use *mierda* rather than the literal translation. Another option commonly used when translating ‘fuck’ is to omit the Spanish taboo word, as illustrated in (4). This strategy has been applied in 13.9% of the occurrences of the word ‘fuck’. Finally, 5.9% of the cases have been translated using transposition, method which involves a change in the grammatical category. This is exemplified in (5), where the noun ‘fuck’ has been translated as the Spanish verb ‘joder’. Contrary to what might be expected, its morphological derivative ‘fucking’ has more occurrences than ‘fuck’. Both English words are used with different purposes in the SL; hence they are also rendered in a different way in the TT, as previously stated. Thus, this derived or inflected word is translated literally in 27.9% of the cases, as occurs in (6). Softening, in turn, is the strategy used in 3.9% of the total amount of cases where this swearword is present in a subtitle, as is the case in example (7). Regarding the change of swearword strategy, this taboo word is translated as a different word but preserving its pejorative meaning in 18.9% of the occurrences, as shown in example (8). The most common translation strategy used to translate ‘fucking’ where it is present in the subtitles compiled in the corpus is omission, as it has been used in 46.0% of the cases. For instance, example (9) has been translated into Spanish omitting the taboo word. Finally, the least frequently used strategy to render this word is transposition, which is used in 3.0% of the occurrences of this swearword in the corpus. One such example of this strategy is the

subtitle in example (10), where ‘fucking’, which is a present participle functioning as an adjective, is translated as a prepositional phrase.

- (1) *Even when I fuck I talk.*
Hablaba mientras follaba.
- (2) *What the fuck you gonna do now?*
¿Qué narices vas a hacer ahora?
- (3) *What the fuck!*
¡Qué mierda!
- (4) *Get him the fuck out of here.*
Sácalo de aquí.
- (5) *What the fuck!*
¡Que te jodan!
- (6) *You’re fucking my best friend.*
Estás follándote a mi mejor amigo.
- (7) *I mean, this dude is fucking bad.*
Este tipo es realmente malo.
- (8) *You fucking piece of shit.*
Hijo de puta.
- (9) *They got the whole fucking killing team here.*
Todo el equipo asesino está aquí.
- (10) *You fucking pricks.*
Huevones de mierda.

6.1.2. General translation strategies used for translating ‘shit’

The frequency of use of these translation strategies considering the translation of the word ‘shit’ can be seen both in the graph and the table below.



Graph 3. Frequency of use of translation strategies for translating ‘shit’

	Literal translation		Softening		Change of swearword		Omission		Trans-position		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
shit	785	42.0	485	25.9	243	13.0	299	16.1	56	3.0	1,868	100

Table 5. General translation strategies used to translate ‘shit.’

Considering the samples of language in use that have been analysed, the strategy of literal translation is the commonest one in the corpus, as 42% of the cases have been translated literally, as in example (11). Softening – illustrated by example (12) – accounts for 26% of the cases in which ‘shit’ appears in a subtitle, whereas in 16% of the total amount of cases, as happens in (14), the taboo word is omitted. Translators have opted to use a different swearword only in 13% of the cases. One of these translations changing the taboo word can be seen in example (13). Finally, as done in example (15), ‘shit’ has been rendered changing its word-class in the TL only in 3% of the subtitles from *OpenSubtitles*.

- (11) *I shit myself*
Me cago viva.
- (12) *She treats you like shit.*
Te trata fatal.
- (13) *I don't need this shit.*
No necesito estas idioteces.
- (14) *You can't prove shit.*
No puede probarlo.
- (15) *Captain, this is going to shit us.*
Capitán, ha sido una caída.

6.2. Variables of the translation of taboo words

A certain taboo word is not always translated in the same way. The choice of translation strategy, in this sense, will depend on different factors. Mainly, the two factors that may influence the translation of taboo words are, on the one hand, the grammatical category of the taboo word; and, on the other hand, the pragmatic factors defining the situation in which the taboo word is used. Nevertheless, the pragmatic factors influencing the rendition of subtitles compiled by the corpus are not going to be analysed as the corpus does not provide enough information to reach conclusions in this respect.

6.2.1. Grammatical category

This section is aimed at discussing each and every grammatical class that the studied taboo words can perform. After this analysis, it would be determined whether the grammatical class of the taboo word may have an effect on its translation or not.

6.2.1.1. Fuck and fucking

A thorough search has been made in the corpus so as to determine the grammatical categories of all the examples of those words. Thanks to a tool that the search interface provides, the times that the searched word and its morphological variants appear in the corpus can be seen. However, the same word may have more than just one grammatical category and that has to be manually analysed.

Table 6 below shows a quantitative result of the frequency in which ‘fuck’ and its derivatives appear in the corpus, ranked from the most frequent to the least one:

Form	Grammatical category + example	Number of occurrences of each word out of 3,776
Fucking	Adverb: <i>Frank, it's fucking infested down here</i> Adjective: <i>Stop the fucking car!</i> Verb: <i>Why are you fuckin' with my friend?</i>	2058 (54.5%)
Fuck(s)	Noun: <i>I could give her a good fuck.</i> Adjective: <i>This fuck clown.</i> Verb: <i>Don't fuck with me.</i> Interjection: <i>Oh, fuck!</i>	1329 (35.1%)
Fucked	Verb: <i>You fucked me.</i> Adjective: <i>We're fucked</i>	241 (6.3%)
(mother)fucker(s)	Noun: <i>This fucker right here.</i> <i>You are a mother-fucker!</i>	62 (2.3%)

Table 6. Frequency of grammatical categories of the word ‘fuck’ and its most common variants.

As can be seen above, the word ‘fucking’ is by far the most used morphological variant in this corpus, with 54.5% of the total amount of occurrences. The popularity of this word is not restricted to this corpus but to the English language in general. This fact was previously noted by Pujol (2006: 124) and Rojo and Valenzuela (2000: 209).

		Literal translation		Softening		Change of swearword		Omission		Trans-position		Total	
		N	%	N	%	N	%	N	%	N	%	N	%
Fuck	N	13	2.3	-	-	319	57.1	186	33.3	40	7.1	558	100
	Adj	-	-	-	-	13	100	-	-	-	-	13	100
	V	346	70.3	80	16.2	27	5.4	-	-	39	7.9	492	100
	Int	173	65.0	13	4.8	80	30.0	-	-	-	-	266	100
Fucking	Adj	370	25.7	41	2.8	391	27.2	576	40.0	62	4.3	1440	100
	Adv	21	5.4	-	-	-	-	371	94.6	-	-	392	100
	V	185	81.9	41	18.1	-	-	-	-	-	-	226	100

Table 7. Translation strategies used in ‘fuck’ and ‘fucking’ considering the grammatical class.

After analysing the different examples of the corpora, some conclusions can be drawn. First of all, the results obtained for ‘fuck’ are going to be illustrated, followed by the results extracted from the study of the translation of the taboo word ‘fucking’. The former taboo word can perform the role of noun, adjective, verb, and interjection as well. When ‘fuck’ as noun is to be translated, the most common procedure is to change the swearword of the TT. Cases where this is carried out like (3) stand for 57.1% of times when this noun appears. Then, in 33.3% of the cases, the omission strategy is used so as to translate ‘fuck’ as noun, as has been done in (16). Example (17) is a case where transposition has been employed. In this case, the taboo word is originally a noun that is translated as a verb. This technique is used in 7.1% of the cases. Last but not least, this noun can also be rendered literally, as in (18). This is done in 2.3% of the cases in which ‘fuck’ works as a noun.

Afterwards, as listed above, ‘fuck’ can also function as an adjective. The only strategy used in our corpus to translate these cases is through change of swearword as example (19) illustrates.

It is quite frequent for ‘fuck’ to work as a verb. When this is the case, its literal translation would be any conjugation of the Spanish verbs ‘joder’ and ‘follar’ as in respectively (1) and (6) above. For ‘fuck’, this happens in 70.3% of the occurrences of this verb. However, it can also be softened and translated as *jugar*, *acostarse*, *copular*, *tener sexo*, *bromear*, *penetrar*, and *hacer el amor*. Example (20) shows how this verb is softened into the first verb of the previous list. Moreover, this verb can be translated as a different one but preserving its taboo content through change of swearword. Examples like (21) where this strategy is used stand for 5.4% of the cases. In addition, in 7.9% of

the occasions on which this verb appears in subtitles, it is translated as a different grammatical class. One such example can be seen in (22), where this verb turns into a noun in the TT.

Finally, it is quite common to find the taboo word ‘fuck’ working as interjection in subtitles. These are as in (23), mainly translated literally, as the frequency stands for 65.0%. In addition, translators have used the strategy of change of swearword in 30.0% of the cases in which this swearword appears as interjection, such as in (24). Lastly, the pejorative meaning can be softened as well, although this strategy is less frequent. Samples like (25), where this strategy is carried out, represent 4.8% of the total amount the examples where ‘fuck’ works as interjection.

Moving forward to the taboo word ‘fucking’, the rendition of this word when it works as adjective is not an easy task, as mentioned by Rojo and Valenzuela (2000: 211). This word may represent the grammatical categories of adjective, adverb and verb. The first grammatical class is the most frequent for this word. In fact, it performs the function of adjective in 1,440 cases out of 2,058. In these cases, it is omitted in 40.0% and translated literally in 25.7% of the times it appears in subtitles, as illustrated by (26) and (27) respectively. This adjective can also be softened, as in (28), although rarely, as this strategy is used only in 2.8% of the cases in which this taboo word is an adjective. Translators can also change the swearword when translating the adjective ‘fucking’ as happens in (29) and in 27.2% of the occurrences. Finally, this adjective can also be translated using transposition, as example (10) shows. It is translated as a different grammatical category in 4.3% of the cases.

After having shown the frequencies for each translation strategy to render ‘fucking’ when it functions as an adjective, it is time now to discuss how this swearword is translated when it works as an adverb. Only two translation strategies are used to accomplish this task in this corpus. One of them is literal translation, when this adverb is translated as *jodidamente*, as in example (30). This strategy is not commonly used; only 5.4% of the cases have been translated in this way. The other strategy, which is the commonest one, accounting for 94.6% of the cases, is omission, illustrated by example (31).

Apart from this, it can be stated that, when working as present participle in a progressive verb form, ‘fucking’ is frequently translated literally, as in (6). This translation strategy stands for 81.9% of the cases in which this taboo word performs the role of verb. Nonetheless, this is not the only strategy used to render ‘fucking’ as verb,

as example (32) demonstrates, since it has been softened in the Spanish subtitles in 18.1% of the cases.

(16) *Let's get the fuck out of here.*

Vámonos de aquí.

(17) *What the fuck.*

Que te jodan.

(18) *This fuck you told me about.*

Sí, el jodido ese del que me hablaste.

(19) *This fuck clown.*

Este puto payaso.

(20) *I'd love to fuck your girl.*

Me encantaría hacérselo a tu chica.

(21) *To fuck a porn star in a cheap motel room?*

¿Tirarte a una estrella porno en un cuarto de motel barato?

(22) *What Caroline needs is to fuck.*

Lo único que necesita es un polvo.

(23) *Oh, fuck!*

¡Joder!

(24) *Oh, fuck!*

¡Mierda!

(25) *Oh, fuck!*

¡Caracoles!

(26) *Now who's the fucking child?*

¿Quién es el niño ahora?

(27) *Frost is trying to... trigger a fucking vampire apocalypse.*

Frost está tratando de... provocar un jodido apocalipsis de vampiros.

(28) *What a fucking tight ass you got!*

Tienes un precioso culo apretado.

(29) *Don't forget your fucking shoes.*

No olvides tus malditos zapatos.

(30) *That was fucking weird.*

Lo que pasó fue jodidamente raro.

(31) *I fucking love this kitchen.*

Me encanta esta cocina.

(32) *He was fucking with me.*

Estaba bromeando.

6.2.1.2. Shit

After a detailed analysis of the different grammatical categories that the word ‘fuck’ and its variants can have, it is time to discuss the same aspect with respect to the word ‘shit’ and its most frequent variations. With this purpose, a table has been made so as to illustrate this information.

Form	Grammatical category + example	Number of occurrences of each word out of 1,959
Shit(s)	Noun: <i>I’m supposed to buy this shit?</i> Verb: <i>You want to shit sparks?</i> Interjection: <i>Holy shit!</i>	1868 (95.3%)
Shitty	Adjective: <i>There’s shitty water all over the floor.</i>	48 (2.4%)
Shitting	Verb: <i>You’ve got to be shitting me.</i>	28 (1.4%)
Shithead	Noun: <i>Come here, shithead.</i>	15 (0.7%)

Table 8. Frequency on which grammatical categories of ‘shit’ and its most common variants appears in *OpenSubtitles*.

As the table above shows, the different variants of the word ‘fuck’ are with no doubt more common than those of ‘shit’ as can be noticed if results are compared. ‘Shit’ appears in 95.3% of the total amount of occurrences of this word and its morphological variants within the corpus used for this research. As a result, this study is focused only on the translation strategies used to translate the word ‘shit’ depending on its word-class.

	Literal translation		Softening		Change of swearword		Omission		Trans-position		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Noun	413	47.0	249	28.3	49	5.5	148	16.8	14	1.5	878	47.1
Verb	37	100	-	-	-	-	-	-	-	-	37	1.9
Interjection	390	48.5	153	19.0	173	21.5	87	10.8	-	-	803	42.9
Adjective	56	37.3	26	17.3	56	37.3	84	56	28	18.6	150	8.1

Table 9. Translation strategies used in ‘shit’ considering the grammatical class.

The results of our analysis have enabled us to reach some conclusions as far as the grammatical category of ‘shit’ is concerned with respect to its translation. First of all, when ‘shit’ has the grammatical category of noun, the taboo word does not always convey the denotative meaning of excrements. Sometimes it is employed as a pejorative synonym of ‘thing,’²⁴ as in (33). It is important to emphasize the fact that the Spanish

²⁴ ‘Fuck’ may imply this meaning as well, but hardly ever.

noun *mierda* can also be used with this purpose, as example (34) demonstrates. We can refer to any object in a scornful way using this taboo word in both languages. In this context, if the translator uses literal translation, the rendering of the word will perfectly fit the context. For this reason, it can be said that ‘shit’ and *mierda* are equivalents in more than one context; they have similar connotative meanings apart from the literal one. That is the reason why, when functioning as noun, the strategy of literal translation is present in 47% of the cases. Moreover, some translations of this word have also been registered, such as *estupidez/ces*, *tontería(s)*, *cosa(s)*, etc.; one such example is shown in (35). This softening strategy is used in 28.3% of the studied cases when it performs the role of noun. Apart from this, this taboo word can be omitted, as in (36). The percentage reached by omission in our corpus is 16.8%.

The word ‘shit’ is frequently used as an interjection in fixed expressions such as ‘Oh, shit!’ or ‘Holy shit!’ In fact, this category stands for 42.9% of the occurrences of this taboo in the corpus. When belonging to this grammatical category, as in (37), the word is translated literally into the Spanish equivalent in 48.5% of the cases. However, in 19% of the occurrences it is softened, so that a less pejorative interjection is used, such as that shown in (38). It is quite common as well to apply the change of swearword strategy to translate ‘shit.’ This strategy is employed in 18.9% of the cases, where translators use different swearwords, such as *demonios*, *maldición*, or *joder*, as in (39).

Finally, it is not so frequent to find this taboo performing the role of verb. Nevertheless, when it does, the literal translation is the only strategy used concerning the samples of the corpus used in this study. It is denotatively translated as *cagar* in 100% of the cases. Example (40) may serve to illustrate the use of the verb ‘shit’ and its translation.

(33) *Now, take my shit.*

Llévate las cosas.

(34) *You can believe that shit!*

¡Puedes creer esa mierda!

(35) *Stop pointing that shit on me!*

¡Deja de apuntarme con esa cosa!

(36) *I tell you, I’m running this shit!*

¡Yo estoy al mando!

(37) *Oh, shit!*

Mierda

(38) *No shit!*

¡No me digas!

(39) *Holy shit!*

¡Qué demonios!

(40) – *I think I just shit my pants.*

– *Creo que me he cagado en los pantalones.*

Considering the obtained results as far as the grammatical class is concerned, it can be stated that it is an important factor when translating taboo words in subtitles, as it influences upon the translation strategy selected to render taboo words. Consequently, as the percentages reached by the different translation strategies across grammatical categories demonstrate, the taboo word is rendered in one way or another depending on the word-class it belongs to.

7. Conclusion

This dissertation has focused on the translation strategies that subtitlers can use when translating taboo words from English into Spanish. The results of the analysis prove that, as previously mentioned and highlighted by Surià (2012), there cannot be an ultimate and perfect rendition of taboo words, but indefinite and subjective versions. This scholar defends that almost everything is relative in translation and even more when dealing with taboo matters. After finishing the present research, it can be determined that, as Rojo and Valenzuela (2000: 208) stated, an accurate translation has to consider the syntactic category of the swearword and the modified word, if any. As mentioned more than once and defended in Rojo and Valenzuela (2000: 208), ignoring all these fundamental aspects would bring inaccurate translations where the English swearword is translated systematically into the most common equivalent. Therefore, the translation of one specific taboo word cannot be generalized, but some patterns of translation can be established. For instance, according to the results obtained by this research, in 43% of the cases, the word is either omitted or softened. This, as believed by Díaz Cintas and Remael (2014: 144), may be caused by the fact that taboo words can be perceived much ruder if written than if uttered. Furthermore, the great frequency of the use of omission may be due to the fact that very often subtitlers have to shorten the text due to the technical constraints previously discussed in chapter 2. The length of a subtitle is limited to 78 characters if consisting of two lines and to 39 characters if consisting of one line. The need to synchronize subtitles and the uttered words to which they correspond may also lead to apply reduction. Hence, if the taboo word is not considered essential from the perspective of denotative meaning, the translator may opt to delete it.

In addition, this research indicates that the grammatical category variable has an effect on the choice of translation strategy. Thus, for instance, when the three taboo words are used as verbs, they are commonly translated literally, especially if they convey their denotative meaning. On the contrary, when ‘fuck’ is a noun, for instance, the most frequent translation strategy is change of swearword.

Finally, cross-linguistic differences between SL (English) and TL (Spanish) are also very relevant in this case. In this sense, for example, as mentioned previously and stated by Rojo and Valenzuela (2000: 211), the rendition of the word ‘fucking’ can be harsh. This is particularly so when its grammatical category is that of adjective or

adverb, owing to the fact that in those cases there are no direct Spanish counterparts which are normally used in Spanish natural discourse. Consequently, translators more often than not resort to the omission in those cases, to such an extent that when ‘fucking’ is an adverb that strategy reaches 94.6%.

It is important to highlight that the present study has some limitations. Thus, for instance, apart from the grammatical category of the taboo word, there may be other factors which could affect the selection of translation strategy. In this sense, the research tool itself has its own limitations, as it makes difficult to determine certain sociopragmatic factors, as it does not offer access to the subtitles in their natural multi-semiotic form, as the *Veiga Corpus* does. Besides, the corpus used we used as research tool does not offer statistics regarding the different counterparts of a given ST word, which implies that for this study all the bilingual units had to be manually analysed and categorised, which was a really time-consuming task.

Bibliographical references

- Biber, D., & Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Díaz Cintas, J., & Remael, A. (2014). *Audiovisual Translation: Subtitling*. New York: Routledge.
- Ferklová, S. (2014). *Subtitles vs. Dubbing: Approaches to Translation of Swear Words and Slang in Film*. Retrieved from Masaryk University Information System. http://is.muni.cz/th/263418/ff_m/?lang=en. Last Access: 17/04/2016
- Han, C., & Wang, K. (2014). Subtitling swearwords in reality TV series from English into Chinese: A corpus-based study of *The Family*. *Translation & Interpreting*, 6(2). DOI: ti.106202.2014.a01
- Jay, T. (2009). The Utility and Ubiquity of Taboo Words. *Perspectives on Psychological Science*, 4(2). Retrieved from https://www.mcla.edu/Assets/uploads/MCLA/import/www.mcla.edu/Undergraduate/uploads/textWidget/1457.00018/documents/Jay_Taboo.pdf. Last Access: 11/05/06
- Jay, T., & Janschewitz, K. (2008). The pragmatics of Swearing. *Journal of Politeness Research*, 4. Retrieved from <http://www.mcla.edu/Assets/MCLA-Files/Academics/Undergraduate/Psychology/Pragmaticsofswearing.pdf>. Last Access: 11/05/16
- Marleau, L. (1982). Les sous-titres... un mal nécessaire. *Meta: Journal Des Traducteurs*, 27(3), 271-851. Retrieved from <https://www.erudit.org/revue/meta/1982/v27/n3/003577ar.pdf>. Last Access: 07/05/16
- Nesselhauf, N. (2005). *Corpus Linguistics: A Practical Introduction*. Retrieved from <http://www.as.uni-heidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf>. Last Access: 10/05/16
- Pérez-Ávila, E. (2006). *El corpus lingüístico en la didáctica del léxico en el aula de E/LE* (Master dissertation). Retrieved from http://www.mecd.gob.es/dctm/redele/Material-RedEle/Biblioteca/2010_BV_11/2010_BV_11_1er_trimestre/2010_BV_11_01Alonso.pdf?documentId=0901e72b80e18f0a. Last Access: 15/05/16
- Pettit, Z. (2009). Connecting Cultures: Cultural Transfer in Subtitling and Dubbing. *New Trends in Audiovisual Translation*, chapter 3.
- Pluszczyk, A. (2015). *An analysis of swearing from a positive perspective* (Doctoral dissertation). Retrieved from International Papers on English Studies. http://www.ipesjournal.org/ipjesjournal/Current_1_files/Ipes%20Adam.pdf Last Access: 17/04/16
- Ponce Márquez, N. (2007). El apasionante mundo del traductor como eslabón invisible entre lenguas y culturas. *Revista Electrónica de Estudios Filológicos*, XIII. Retrieved from https://www.um.es/tonosdigital/znum13/secciones/tritonos_B_nuria%20Ponce.htm. Last Access: 06/04/16
- Pujol, D. (2006). The Translation and Dubbing of “Fuck” into Catalan: The Case of *From Dusk till Dawn*. *The Journal of Specialised Translation*. Retrieved from http://www.jostrans.org/issue06/art_pujol.php. Last Access: 29/04/16
- Rica Peromingo, J. (2014). La traducción de marcadores discursivos (DM) inglés-español en los subtítulos de películas: un estudio de corpus. *The Journal of Specialised Translation*, 21, 177-199. Retrieved from http://www.jostrans.org/issue21/art_rica_peromingo.pdf. Last Access: 10/05/16

- Rica Peromingo, J., & Albarrán, R., & García, B. (2014). New Approaches to Audiovisual Translation: The Usefulness of Corpus-Based Studies for the Teaching of Dubbing and Subtitling. *Languages for Specific Purposes in the Digital Era* (303-321). Switzerland: Springer International Publishing Switzerland.
- Romero Fresco, P. (2005). *The translation of phraseology in a parallel (English-Spanish) audiovisual corpus. The case of Friends* (Doctoral Dissertation). Retrieved from University of Birmingham
<http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>. Last Access: 06/04/16
- Rojo, A. M., & Valenzuela, J. (2000). Sobre la traducción de las palabras tabú. *Revista de Investigación Lingüística*, 3(1), 207-220
<https://digitum.um.es/jspui/bitstream/10201/16355/1/4231.pdf>. Last Access: 29/04/16
- Sotelo, P. (2012). Using a Multimedia Parallel Corpus to Investigate English-Galician Subtitling. Retrieved from
<http://sli.uvigo.es/arquivos/sdh2011.pdf>. Last Access: 10/04/16
- Surià, S. (2012). En la luna de Babel: *El traductor insolente: traducir el lenguaje soez* [Blog sobre lenguas y traducción]. Retrieved from
<http://enlalunadebabel.com/2012/12/10/el-traductor-insolente-traducir-el-lenguaje-soez/>. Last Access: 15/05/16
- Tiedemann, J. (2007). *Building a multilingual parallel subtitle corpus*. (Doctoral dissertation). Retrieved from
<http://stp.lingfil.uu.se/~joerg/paper/clin17.pdf>. Last Access: 10/05/16
- Tiedemann, J. (2009). *News from OPUS – A collection of multilingual parallel corpora with tools and interfaces* (237-248). Retrieved from
<http://stp.lingfil.uu.se/~joerg/published/ranlp-V.pdf>. Last Access: 10/04/16
- Tiedemann, J. (2010). *Parallel Data, Tools and Interfaces in OPUS*. (Doctoral dissertation). Retrieved from
http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf. Last Access: 11/05/16
- Wang, N. (2013). An analysis of the pragmatic functions of “swearing” in interpersonal talk. *Griffith Working Papers in Pragmatics and Intercultural Communication*, 6. Retrieved from https://www.griffith.edu.au/data/assets/pdf_file/0007/589453/Na-Wang.pdf. Last Access: 17/04/16
- Xiao, H., & Wang, X. (2009). Constructing Parallel Corpus from Movie Subtitles. In *Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy* (329-336). Retrieved from
http://home.in.tum.de/~xiaoh/pub/iccpol_09.pdf. Last Access: 10/05/16