Subtitling Swear Words from English into Chinese: A Corpus-Based Study of Big Little Lies

Shuangjiao Wu

Zhejiang Yuexiu University, Shaoxing, China
Email: sky_shuangjiao@163.com

Abstract

As Jay and Janschewitz (2008) highlight, the pragmatic function of swear words is to express emotions, such as anger and frustration. The main objective of the present paper is to analyze the translation of the two commonest swear words in English, namely fuck and shit (Jay, 2009: p. 156; Rojo & Valenzuela, 2000: p. 209) and their morphological variants into Chinese in subtitles of the TV series Big Little Lies Season 1. The research instrument used in this study has been a parallel corpus with the software Paraconc, a bilingual English-Chinese corpus of subtitles. Regarding the research results, it could be observed in the findings of translation solutions, among which 66.3% of the instances were rendered in a sanitized manner, including omission taking 31.6%, softening 27.6% and de-swearing 7.1%. In addition, the factors such as the variables of swear words and the grammatical category of swear words also have an influence on the selection of translation solutions.

Keywords

Subtitling, Swear Words, Translation Solution, Corpus-Based Study

1. Introduction

The presence of swear words in subtitles is a translation problem which has lately received scholars’ increasing attention. The main pragmatic function of these lexical items, as Jay and Janschewitz (2008) highlight, is to express emotions, such as anger and frustration (Díaz-Pérez, 2020). As Greenall (2011) states following Mao (1996), swearing generates social implicature, which means that “it gives valuable hints regarding aspects of individuality and class membership, information which is crucial in understanding where someone comes from” (Greenall, 2011: p. 45).

The main objective of the present paper is to analyze the translation of the two commonest swear words in English, namely fuck and shit (Jay, 2009: p. 156; Rojo & Valenzuela, 2000: p. 209) and their morphological variants into Chinese in subtitles of the TV series Big Little Lies Season 1. The research instrument used in this study has been a parallel corpus with the software Paraconc, a bilingual English-Chinese corpus of subtitles. Regarding the research results, it could be observed in the findings of translation solutions, among which 66.3% of the instances were rendered in a sanitized manner, including omission taking 31.6%, softening 27.6% and de-swearing 7.1%. In addition, the factors such as the variables of swear words and the grammatical category of swear words also have an influence on the selection of translation solutions.
jo & Valenzuela, 2000: p. 209) and their morphological variants into Chinese in subtitles of the TV series Big Little Lies Season 1. The research instrument used in this study has been a parallel corpus with the software Paraconc. It’s a bilingual English-Chinese corpus of subtitles. Considering the total number of instances of the two swear words and morphological variants retrieved from the corpus, 98 examples have been analyzed. Two variables have been investigated to check whether there was interdependence between each of them and the choice of translation solution. Those two variables were the specific swear word—fuck v.s. shit and the grammatical category or part of speech: noun, verb, adjective, adverb, and interjection.

The translation of swear words in subtitling is an under-researched area in audiovisual translation in general and in subtitling in particular, as emphasized by Cabrera & Javier (2016a: p. 38), who says that the use of taboo language was more researched in dubbing than in subtitling (Díaz-Pérez, 2020). While a growing interest in AVT research can be seen in many European countries, little has been done in the Chinese world, where to the best of the researcher’s knowledge, very few studies have been done to investigate the different translation solutions of swear words, and whether the variables of swear words could have an effect on those translation solutions used in their translations in the Chinese culture. Therefore, this paper aims to contribute to this academic field.

In the following, before concentrating on different methodological aspects in Section 3, such as the description of the corpus and the different research stages, Section 2 will deal with the translation of swear words. Afterwards, in Section 4, the results of the study will be presented and discussed. The article will close with some concluding remarks contained in Section 5.

2. On the Translation of Swear Words

It has been stated (Rojo & Valenzuela, 2000; Fernández Dobao, 2006) that the translation of swear words is a delicate issue and that cross-cultural differences regarding swearing should be taken into account by the translator. As put forward on several occasions (Díaz Cintas, 2001; Chen, 2004; Hjort, 2009; Díaz Cintas & Remael, 2014; Han & Wang, 2014; Díaz-Pérez, 2020), swearing and taboo words tend to be toned down due to several reasons.

Amongst the reasons which might explain the resort to the omission solution and to the sanitizing tendency in general, it has been highlighted (Mayoral, 1993; Ivarsson & Carroll, 1998; Díaz Cintas, 2001; Chen, 2004; Han & Wang, 2014; Cabrera & Javier, 2015, 2016a, 2016b; Santamaría Ciordia, 2016; Díaz-Pérez, 2020), for instance, that the impact of swear words in the written language is reinforced as compared to their use in oral speech.

In this sense, Hjort (2009) reports that in a questionnaire she administered to translators, 93% of the informants agreed that swear words were stronger when written in subtitles than when uttered in oral speech (Díaz-Pérez, 2020). Díaz Cintas (2001: p. 51) argues that “the mere fact of the graphic and material repre-
sentation is not a sufficient enough reason to justify this discrepancy in value and that a crucial element has been missed in this debate.” In addition, Francisco Javier Díaz-Pérez (2020) highlights that “the context in which reading takes place must also be taken into account according to Díaz Cintas, since it is not the same to read a novel in private as to read the subtitles of a film in public. Although in the end reading is always an individual act, a high amount of taboo words in the subtitles of a film may be perceived as more aggressive than the same quantity of taboo words in a novel”. Therefore, it’s one of the reasons that swear words, in some cases, are omitted in subtitling.

The fact that the lack of a direct counterpart of the source text (ST) swear word in the target language (TL) is another reason which may explain both the omission of swear words and the sanitizing tendency in the target text (TT) (Díaz-Pérez, 2020). Swearing is culture-specific, and as highlighted by Han and Wang (2014: p. 1), literal translations of ST swear words with no existing direct counterparts in the TL would most likely be considered unnatural by the TT viewer.

The audiovisual translation mode with which we are concerned in this study is subject to certain technical restrictions which may also have some consequences for the translation of swear words. It is known that subtitling is a specialised translation. It involves not only the transfer of one language into another language, but also the matching of the subtitles with the spoken word and the visual images on the screen. The unique characteristics of subtitling add technical constraints for subtitlers. For example, the display of a chunk of subtitles only allows a minimum of 1.5 to 2 seconds and a maximum of 6.5 to 7 seconds on the screen (Gottlieb, 1998: p. 1008). Essential visual information might be blocked if more than two lines and more than 34 - 37 characters per line are displayed on a screen (ibid: 1008). In addition, the onset of a subtitle needs to be aligned with the camera change as much as possible to create a synchronous viewing experience for the audience.

Thus, the number of characters per line and per subtitle is limited, as is the reading speed which may be demanded from the viewer. Related to this consideration, the fact that swear words do not convey denotational meaning which may be considered essential for the development of the plot in a film may lead the subtitler to get rid of swear words in case of necessity. Due to the very nature of audiovisual texts, viewers may rely on both the audio and visual channels to retrieve some information. Prosodic features, such as stress, intonation, or loudness, and kinesic features, such as body and face gestures, may provide the viewer with an invaluable help to interpret a given expression as an emotionally charged expletive (Díaz-Pérez, 2020).

3. Methodology
3.1. Research Objectives

The general objective of the present study, as described above, is to analyze the
translation of the commonest English swearwords—fuck and shit with their morphologic variants in the parallel corpus, an English-Chinese corpus of subtitles of the TV series Big Little Lies Season 1.

In order to attain this general objective, the following specific aims have been pursued:
- To retrieve and classify all the instances of fuck and shit together with their morphological variants in the English sub-corpus;
- To classify all the translation solutions adopted to translate the SL swearwords in the Chinese subtitles;
- To determine whether the swearword variable—fuck vs shit has an effect on the choice of translation solution;
- To analyze whether the grammatical category and the translation solution variables are related or independent.

3.2. Data Collection

Hatim and Mason (1997: p. 70) suggest that the audiovisual work that suits research should 1) be a widely-distributed, full-length feature work with high quality subtitles; 2) be work where interpersonal pragmatics are brought to the fore; and 3) contain many sequences of verbal interaction such as sparring. (Cheng, 2019).

The corpus being used in this research is the first season of Big Little Lies produced by HBO, which contains seven episodes. Based on Liane Moriarty’s best seller and featuring Reese Witherspoon, Nicole Kidman, Shailene Woodley and more. Big Little Lies is a dark comedy set in a town by the seaside in California. The TV series Big Little Lies made itself a great success due to the acting and scripts. It was rated highly on different film commentary websites, with ratings such as 9.0/10 (Douban Movies, 2017), 89% (Rotten Tomatoes, 2017) and 8.5/10 (IMDb, 2017). Thus, the TV series Big Little Lies meets the above standard set by Hatim and Mason (1997), and fits the purpose of this research.

The Chinese translation chosen in this article is done by “YYeTs” (人人影视, ren-ren-ying-shi), being one of the largest fansub groups in China and known for its reputable quality in audiovisual translation. Fansub (a short form of “fan subtitled”) is an emerging topic in Translation Studies. It was earlier defined by Diaz Cintas as “a fan-produced, translated, subtitled version of a Japanese anime programme”, but now also refers to such versions of other audiovisual works (Cheng, 2019).

Because of its high quality, YYeTs’s translation has been used by various legitimate video distributors online, including Souhu (搜狐, sou-hu), Youku (优酷, you-ku) and 163.com (网易, wang-yi) (Wang & Zhang, 2017). Such recognition gives credence to the translation quality of Big Little Lies carried out by YYeTs. This research uses the proofread version of the translation to build the corpus.
To build the corpus, both English and Chinese subtitles from *Big Little Lies* from season one on the YYeTs website were collected. All seven episodes of season one are selected based on the translation quality. The English and Chinese subtitles are paralleled manually, because the number of the two versions of subtitles doesn’t match. The reason is that there is copyright information about translation done by the fansub YYeTs. Thus, such information has been deleted in the process of paralleling.

Then, a corpus is established consisting of three columns after paralleling: the serial number of the subtitle, the source text, and the target text. A snapshot of the corpus from *Big Little Lies* Season 1 episode 1 is like this:

![Snapshot of the corpus](image)

### 3.3. Research Stages

The different steps or stages followed in the experimental part of this research study can be briefly summarized as follows:

- Retrieval of all the instances of *fuck* and *shit* from the parallel corpus;
- Classification of all the examples regarding grammatical category;
- Identification and clarification of the translation solution adopted in each case;
- Analysis of the results and drawing of conclusions.

### 4. Results and Discussion

The results of the study are presented in this section. Thus Sub-section 4.1 focuses on the presence of *fuck* and *shit* in the English sub-corpus, indicating the frequency of the different morphologic variants in each of the cases. The next Section—4.2 deals with the translation solutions adopted in the whole corpus. Finally, in Section 4.3, analysis will be carried out to determine whether the swear word variables were related to the translation solution.
4.1. *Fuck* and *Shit* in the ST Corpus

As Figure 1 and Figure 2 reflect, the most frequent swear form in the English subtitles is *fucking*, accounting for 54.1% of the total number of the ST swear words analyzed, followed by *shit* used as a noun, which represents 13.3%, and *fuck* as a verb, with 11.2%. These results confirm the findings of other studies which also established that *fucking* was the commonest swear word, such as Rojo and Valenzuela (2000), Leech, Rayson and Wilson (2001), or Fernández & Jesús (2009) and Díaz-Pérez (2020).

Of all the 84 instances of *fuck* in the corpus, Table 1 and Table 2 in the following reflect the frequency of forms or morphological variants of *fuck* and *shit* respectively.

![Figure 1. Frequency of swear word forms in the whole corpus.](image1)

![Figure 2. Percentage of swear word forms in the whole corpus (%).](image2)
Table 1. Forms of fuck in the corpus.

<table>
<thead>
<tr>
<th>From</th>
<th>From N</th>
<th>%</th>
<th>To N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>fucking</td>
<td>37</td>
<td>44.05</td>
<td>53</td>
<td>63.10</td>
</tr>
<tr>
<td>fucking(Adj.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fucking(Adv.)</td>
<td>16</td>
<td>19.05</td>
<td>11</td>
<td>13.10</td>
</tr>
<tr>
<td>fuck(V)</td>
<td>11</td>
<td>13.10</td>
<td>9</td>
<td>10.71</td>
</tr>
<tr>
<td>fuck(Interjection)</td>
<td>9</td>
<td>10.71</td>
<td>9</td>
<td>10.71</td>
</tr>
<tr>
<td>fucked(Adj.)</td>
<td>1</td>
<td>1.19</td>
<td>4</td>
<td>4.76</td>
</tr>
<tr>
<td>fucked-up</td>
<td>3</td>
<td>3.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fucker</td>
<td>2</td>
<td>2.38</td>
<td>2</td>
<td>2.38</td>
</tr>
<tr>
<td>fuck(N)</td>
<td>3</td>
<td>3.57</td>
<td>5</td>
<td>5.95</td>
</tr>
<tr>
<td>a fuck-up</td>
<td>2</td>
<td>2.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>84</strong></td>
<td><strong>100</strong></td>
<td><strong>84</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 2. Forms of shit in the corpus.

<table>
<thead>
<tr>
<th>From</th>
<th>From N</th>
<th>%</th>
<th>To N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shit(N)</td>
<td>10</td>
<td>71.43</td>
<td>13</td>
<td>92.86</td>
</tr>
<tr>
<td>Beat the shit</td>
<td>2</td>
<td>14.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full of shit</td>
<td>1</td>
<td>7.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bullshit</td>
<td>1</td>
<td>7.14</td>
<td>1</td>
<td>7.14</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14</strong></td>
<td><strong>100.00</strong></td>
<td><strong>14</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

4.2. Translation Solutions

As regards the translation of the ST swear words, the translation solutions identified in this corpus have been the following ones:

- Pragmatic equivalence;
- Softening;
- De-swearing;
- Omission.

By using pragmatic equivalence, the author refers to the use of swear word in TT that could realize the functional equivalence in both of tone and pragmatic function (Díaz-Pérez, 2020). Such TT may or may not be the literal translation of the swear word in ST. Thus, in Example 1, the swear word fuck is translated into 操 (cao) in the TT, which is very close in the tone and pragmatic function as that in the ST.

Example 1:

**English Subtitles [Fuck].**

**Chinese Subtitles:** 操

**Back Translation:** Fuck.

However, in Example 2, although the fuck has not literally translated into the TL, the rendition into the mother-theme swearing interjection 他妈的
(ta-ma-de, his mother’s) would give rise to a strong implicature of irritation, since it is a natural and typically vulgar way for Chinese speakers to express anger and typically vulgar way for Chinese speakers to express anger and annoyance.

Example 2:

English Subtitles: But, you know, [fuck] convention.
Chinese Subtitles: 走他妈的清规戒律
Back Translation: Fuck his mother’s convention.

The label softening is used here to refer to the translation of the ST taboo word by means of a TL swear word with a softer or milder tone (Díaz-Pérez, 2020). An example of this solution is using the word 人渣 (ren-zha, human scum) in the TT to refer to huge fuck-up in the ST. In this case, the rendition in TT is much milder in tone than its counterpart in ST.

Example 3:

English Subtitles: Rich, powerful, but DNA-huge [fuck-up].
Chinese Subtitles: 富得流油 权势惊人 但本质上就是个人渣
Back Translation: Rich and powerful, but a huge “human scum”.

The following Example 4 and 5 are also instances of softening, since the Chinese rendition 混蛋 (hun-dan, bastard) in Example 4 and 混犊子 (hun-du-zi, bastard) in Example 5 are both taboo words in TT, which contain the meaning of swear word, but are both much milder in tone.

Example 4:

English Subtitles: Instead of some little [fucker] who’s chomping on our kid?
Chinese Subtitles: 而不是欺负咱家宝贝女儿的混蛋吗
Back Translation: Instead of being a bastard who’s chomping on our kid?

Example 5:

English Subtitles: … you fucking little [shit].
Chinese Subtitles: 你个 混犊子
Back Translation: … you such as bastard.

The third solution is termed de-swearing, as the name indicates, which refers to the translation of English swearwords into plain, non-swearwords in Chinese. However, the meaning of swearing contained in the ST is delivered in the TT. The following examples are instances of de-swearing. In Example 6, the Chinese rendition 脏手 (zang-shou, dirty hand) is used to express the meaning of swearing of “fucking hand”. In addition, the swear word fucked-up ideas in the ST is translated as the Chinese word 馊主意 (sou-zhu-yi, bad idea) in the TT.
Example 6:

English Subtitles: I asked you to remove your [fucking] hand.
Chinese Subtitles: 我叫你把你的脏手拿开
Back Translation: I asked you to remove your dirty hand.

Example 7:

English Subtitles: … is your willingness to go along with your wife’s [fucked-up] ideas.
Chinese Subtitles: 不管妻子有什么馊主意都得配合
Back Translation: … is your willingness to go along with your wife’s bad ideas.

Omission: As the term suggests, this strategy occurs when subtitlers/translators decide to remove a word that is deemed to be seriously offensive or face threatening. The following Example 8 and Example 9 could be used to explain the translation solution. In the TT, no swearing words are used to render the meaning of swearing in the ST.

Example 8:

English Subtitles: … but peel one [fucking] potato with Bonnie.
Chinese Subtitles: 但她跟邦妮削了一个土豆就…
Back Translation: … but peel one potato with Bonnnie.

Example 9:

English Subtitles: Should’ve told the boys to pick all this [shit] up.
Chinese Subtitles: 你该叫孩子们把玩具收好
Back Translation: Should’ve told the boys to pick all this up.

In the whole corpus, the most frequently used translation solution, as shown in Figure 3, is pragmatic equivalence, accounting for 33.7%, which is followed by omission (31.6%) and softening (27.6%). However, if the percentage of omission is added to the softening and de-swearing percentages, as the results shown in Figure 4, about 66.3% of the cases are sanitized when translating the subtitles into Chinese.

This high degree of sanitization confirms the results of other studies, such as Díaz Cintas (2001), Chen (2004), Hjort (2009), Greenall (2001), Han and Wang (2014), or Santamaría Ciordia (2016), and Díaz-Pérez (2020).

Several reasons have been given to explain the above tendency:

First, the linguistic differences between English and Chinese may prevent the translators from rendering one-to-one translation.

Second, as the technical constraints together with the linguistic constraints in the process of subtitling increase the translation challenges and require different translation strategies (Diaz Cintas & Anderman, 2009). The most important
strategy of subtitling, as argued by many researchers (e.g. Diaz Cintas & Anderman, 2009; Georgakopoulou, 2009; Pettit, 2009), is condensation. This strategy aims to convey the plot-carrying message by avoiding verbal redundancies, changing or even omitting such non-critical elements as fillers and exclamations and re-shaping the original linguistic structure. Therefore, due to the limitations in temporal and spatial dimensions, the TT is usually a reduced form of ST. Therefore, the subtitlers are left to deciding which information to retain in TT.

4.3. Variables

4.3.1. Swear Word: *Fuck* and *Shit*

Regarding the translation solutions used to render *fuck* and *shit* respectively, an obvious difference could be detected as shown in Figure 5. The most commonly used solution to translate *fuck* is Pragmatic equivalence, accounting for 32.7%, followed by those of omission and softening, with 24.5%. On the contrary, the most frequent translation strategy in the case of *shit* is omission, about 7.1%, which is followed by those of softening and de-swearing, with 3.1% of the total number of instances of *shit* in the whole corpus.

In spite of a general tendency of sanitizing both for translating the two swear
words into Chinese, the differences in the translation solutions could be detected in the cases of *fuck* and *shit* in the whole corpus. In other words, the variables of swear words have an impact on the translation solutions the subtitlers would use in their translations.

4.3.2. Grammatical Category

**Figure 6** below shows that the translation solutions across grammatical category. The commonest translation solution used for adjectives is omission, accounting for 54.1% in the case of adjectives. However, the most frequent translation strategy in the case of verbs is de-swearing, with 72.2%. The most commonly adopted translation solutions in the cases of adverbs and interjections are de-swearing and softening respectively. In other words, when the swear words are adjectives, they are often omitted in subtitling. In addition, much milder renditions are used in Chinese when the English swear words are verbs.

The linguistic differences between English and Chinese might explain the above finding. In general, there isn’t a Chinese adverb equivalent for *fucking*, which could be used as an intensifier of an adjective or a verb, thus pragmatic
equivalence is not commonly used, while the translation solutions of softening and de-swearing are adopted. In addition, although there are Chinese equivalents of *fuck* as a verb, a noun and an interjection, there are far less expressions of swear words than those in English. In addition, it’s believed that subtitlers would face moral and social restraints when working with such cultural taboos as swearwords. They may adopt self-censorship consciously or unconsciously, causing them to omit some of the swearwords, or de-swear some of them in their translations to meet the assumed tolerance and expectation of the target audience.

5. Concluding Remarks

As it is proved throughout this paper, there is a tendency of sanitizing in the Chinese translations of swear words. It could be observed in the findings of translation solutions, among which 66.3% of the instances were rendered in a sanitized manner, including omission taking 31.6%, softening 27.6% and de-swearing 7.1%. In addition, the factors such as the variables of swear words and the grammatical category of swear words have an influence on the selection of translation solutions.

As the above analysis shows, the swear words in this TV series have been toned down in a general perspective. However, the function of swear words has been largely retained in the Chinese subtitles by using the translation solutions such as softening (27.6%) and pragmatic equivalence (33.7%). Even the swearing in English subtitles has been replaced by the less vulgar rendition in Chinese subtitles, the pragmatic function has not largely affected, and the audience could still get a general understanding of the original subtitles.

Therefore, the implication of this study tends to suggest that the focus of subtitles translation should not be given to the omission of swearing, but the translation solutions used to render the swear words, and how the pragmatic function could be retained in the TL in spite of the cultural constraints and the temporal and special constraints in audiovisual translation.

However, the present study is limited in scope, in that no objective measuring of audience’s reception was undertaken. Further studies could draw on other disciplines to investigate these aspects of subtitling.

Acknowledgements

The author wishes to acknowledge support from the Project of Blended Teaching Reform of Zhejiang Yuexiu University (No. JGH1901), and the Project of Higher Education Research of Zhejiang Province Association for Higher Education (No. KT2020136).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.
References


