International Journal of English Language Studies (IJELS)
ISSN: 2707-7578
Website: www.ijels.one

Original Research Article

Shift of Oral and Written Language Features in Audio-visual Translation: A Case study of Subtitling and dubbing of the American TV show 'Prison Break'

Salah Ben Hammou
Doctorate student at Hassan II university, School of Arts and Humanities, Mohammadia, Morocco
Corresponding Author: Salah Ben Hammou, E-mail: salahbenhammou937@gmail.com

ARTICLE INFO

Article History
Received: March 28, 2020
Accepted: June 12, 2020
Volume: 2
Issue: 2

KEYWORDS
Audio-visual Translation, written linguistic features, oral language

ABSTRACT

This study investigates the use of spoken and written linguistic features in audio-visual translation. This kind of translation, also called screen translation, is different from classical types of translation, such as literary translation, technical translation and legal translation, in that the source text is spoken. This oral text can be transferred into a written text, the case of subtitling, or into an oral target text, the case of dubbing. Thus, in order to investigate the characteristics of spoken and written language in audio-visual translation, I have conducted a quantitative comparison between a dubbed and a subtitled version of an American TV Show. I have focused in my study on the choice of oral and written linguistic features in both versions.

Introduction

In linguistics, we generally make a distinction between two modes of language: spoken and written. Each mode has specific characteristics, which distinguish it from the other mode. In other words, the features that characterize the spoken language are different from those that characterize the written language. The nature of this distinction has received much attention from scholars in the early 1970s. Among the early linguists who have studied the differences between spoken and written language: *Goody and Ian watt (1968), Josef Vachek (1973), and Band Oslan (1977)***.** The writing of these linguists marked the end of a period in which the systematic study of language was dominated by *Leonard Bloomfield*'s assumption that writing is not a language, but merely a way of recording by means of visible marks (Bloomfield 1933: 21).

In modern linguistics, the interest in the relationship between spoken and written language continued with several influential linguists; *Tannen (1982), Och (1979) and Chafe (1982-1987)*. These linguists have tried to investigate the nature of differences between spoken and written language in a specific manner without claiming the primacy of one mode over the other. They have worked on data from different genres of discourse to study the properties of both spoken and written discourse. *Chafe***,** for instance, tried to identify the differences in the kinds of linguistic expressions, which are used, by speakers and writers, suggesting that there are underlying causes for those differences. For him, neither spoken nor written language is a monolithic phenomenon. Although each mode allows a multiplicity of styles, they do share many features. This means that some genres of spoken discourse have features that are specific to written discourse (e.g.: *lectures).* Similarly, some types of written discourse (e.g.: *personal letters)* have oral features. (Chafe & Danielewicz 1987:01)

Chafe (1987:02) identified several features that are held to be responsible for the differences in the nature of language a person can use. One of these features is whether language is produced with the mouth and received with the ear, or whether it is produced with the hand and received with the eyes. Other factors are related to the context of language use, the purpose of the speakers or writers and the subject matter of what is being said or written.

Despite this introduction about spoken and written language, my primary interest in this research is not to draw a comparison between spoken and written language and list a number of differences that distinguish between the two types. For this has been the subject of many previous studies by influential linguists like the ones we mentioned above. Instead, I

intend to be more specific and study the use of spoken and written linguistic features in audio-visual translation. This kind of translation, also called *screen translation,* is different from classical types of translation, such as *literary translation, technical translation and legal translation,* in that the source text is spoken. This oral text can be transferred into a written text, the case of subtitling, or into an oral target text, the case of dubbing. Thus, in order to investigate the characteristics of spoken and written language in audio-visual translation, I will conduct a quantitative comparison between a dubbed and a subtitled version of an American TV Show. I will focus in my study on the choice of oral and written linguistic features in both versions.

The data

I have chosen, as a corpus for my study, the original transcript of the American TV Show "Prison Break", season 4, episode 1, and its French dubbed and subtitled versions. "Prison break" is a drama television series created by **Paul Schering.** It was first broadcast by the Fox Broadcasting Company on August 29, 2005. The series revolves around two brothers. One has been arrested and sentenced to death for a crime he did not commit. The other brother, **Michael Scofield**, devises an elaborate plan to help his brother escape from prison. The TV Show consists of four seasons, each season with 22 episodes. Our data is taken from the fourth season, 1st episode. In this season, which is entitled "*Final Prison Break?*" , a story is being told about the events that took place in the previous season and the strange scar on *Sara's* shoulder. *Sara* is arrested and put into prison in Miami Dade Penitentiary. With the general and T-bag in the adjunct men's prison, the general wants *Sara* dead and offers a 100.000 bounty. Michael hears of the bounty and devises a plan to break her out of the prison. At the end, Michael sacrifices himself for his brother, wife and child. Michael Scofield has been played by Neut Worth Miller, while Lincoln Burrows has been played by Dominic Purrell. The series has been distributed all over the world through two versions: the dubbed and subtitled versions. I have chosen this series as a corpus for its success all over the world. According to *the New York Times, "Prison Break" was "most intriguing than most of the new network series and it certainly is one of the most original".* The series averaged 92 million viewers per week in its first season.

Thus, our corpus will consist of three samples. The first sample contains the transcript of the original version: 1st episode: 1st twelve minutes. The second sample contains the French dubs, while the third sample contains the French subtitles. The three samples were taken from a fan club, *www.subscene.com.* This club makes transcripts and subtitles of TV- Shows available for educational and entertainment purposes. This fan club has been chosen because of the quality of the transcripts and subtitles. For instance transcripts were not only fairly accurate but also extremely detailed, including several features that scripts are not likely to present: such as hesitations (e.g., un, er, uh), pauses (e.g., _) repeats(e.g.; I'm I'm, thank god, thank god) and contractions (e.g., you've). The following table presents some general counts of our data:

The version	Number of sentences	Number of clauses	Numbers of words	Average of words per sentence
Original	187	314	1152	6.160
Dubbed	178	267	1183	6.646
Subtitled	181	248	1134	6.265

Table1: general quantitative counts of the data.

Research questions

As is mentioned earlier, my primary interest in this empirical research is to study the use of spoken and written features in audio-visual translation, particularly in dubbing and subtitling. To achieve this goal, I will a conduct a quantitative comparison between the three samples of our corpus. The dubbed and subtitled versions represent the same source spoken language. However, the linguistic features used in both versions are not likely to be the same. This is due mainly to the nature of language being used in each version. The language used in the dubbed version is spoken. Therefore, the dubber is supposed to opt for oral features in his translation. Conversely, the language used in the subtitled version is written. As a result, we expect the subtitler to opt for features that characterize written register.

Investigating this issue, I will try through this case study to provide answers for the following questions:

Main questions:

- ✓ How much different is the subtitled version from the dubbed version in the use of oral and written linguistic features (in other word,» involved" vs. "informational features) ?
- ✓ Which version does reflect the linguistic features used in the original version?

Sub-questions:

- ✓ What are the factors responsible for the choice of particular linguistic features in each version?
- ✓ To what extent can we consider the language used in the dubbed version as "involved" and that used in the subtitled version as "informational"?

Methodology

In this case study, I intend to conduct a quantitative comparison between the original, dubbed and subtitled versions of "prison break", season 4, episode 1. For this purpose, I will apply Biber's Multidimensional Analysis (1988). **MD** is a quantitative corpus-based technique whose primary objective is to find and interpret the co-occurrence of certain linguistic features in a corpus.

In his study of register varieties in English, Biber (1988) used the Multidimensional Analysis in order to prove that different registers, e.g. *'face to face conversation, fiction, academic prose'* reveal different co-occurrence patterns of linguistic features and that these co-occurrence patterns would reflect the major communicative functions of those registers. Biber followed a series of steps. He started with the choice of certain linguistic features that are associated with different functions of the language. For instance, first- and second- person pronouns, contractions and demonstrative pronouns are associated with interactive discourse (spontaneous spoken language), while passive constructions and nominalizations are associated with more formal registers **e.g.:** *academic writings.* By the end, **Biber** identified 5 major dimensions of English through the interpretation of "*the communicative functions most widely shared by the set of co-occurring features defining each dimension*". (Biber ,Conrad, Reppen 1998; p:149).

The 5 dimensions are presented in the following order: involved versus informational production, narrative versus non-narrative discourse, elaborated versus situation dependent reference, overt expression of argumentation, and impersonal versus non-impersonal style. Each dimension is defined by a number of linguistic features which are classified into positive and negative features. I will conduct my case study on Dimension 1: *involved* vs *informational production.* With this dimension, the following features: first-and second- person pronouns, private verbs (e.g., think, believe), contractions, and informal lexical items are held to be characteristic of involved registers. They are associated with interactive texts (e.g., face to face conversation). On the other hand, features such as nouns, nominalizations, prepositions and attributive adjectives are associated with formal texts (e.g., *Academic Writings*). The former features are called positive features while the latter are called the negative features. The positive and negative features are in complementary distribution. In other words, registers that are characterized by high frequencies of the positive features tend to have very low frequencies of the negative features and vice versa. Features such as these are the object of analysis of this case study.

Using **Biber's** dimension 1: involved vs informational, I will measure the frequency of certain linguistic features (positive and negative features) in the three samples chosen for this case study. Since the corpora are of different sizes, the raw frequencies of linguistic features will be normed to a common base to allow for accurate comparison. For instance, both corpora of the dubbed and subtitled versions yielded 3 occurrences of formal vocabulary. However, the dubbed version corpus contains 1183 words, while the subtitled version corpus contains 1134 words. To norm these counts to a rate of occurrences (**e.g.:** frequency per 1.000.000), the raw frequency of the feature will be divided by the total number of words in the corpus and then multiplied by 1.000.000.

(Raw frequency/total numbers of words in the corpus) x base
Thus, for the dubbed version:
(3/1183) x 1.000.000 = 2535, 92
For the subtitled version:
(3/1134) x 1.000.000 = 2645, 50

After normalisation, we see that the dubbed version corpus produced 2535, 92 occurrences of formal lexical items per million words, while the subtitled version yielded 2645, 50 instances of formal vocabulary. This shows that the initial impression that both corpora have the same frequency of formal words (3 instances in each corpus) did not prove to be true. The subtitled version presents a slightly higher frequency of this feature. The results will be presented in both tables and graphic forms.

Results

TYPE / TOKEN RATIO

Generally, genres of spoken language are more limited in variety than those of written language. To measure the variety of lexical items in our samples, we use the mechanical measure: type/token ratio. With this measure, we count the number of different words in a sample (type words) and we divide them by the total number of words in that sample (token words). As a simple illustration, let's first measure the type/token ratio of the following sentence:

"It *can be seen that this phenomenon too is almost equally present in both styles of spoken language, whereas its frequency is considerably lower in either styles of writing."*

This sentence contains 29 token words and 25 type words. The words "is", "in", "styles" occur twice. Thus, *the type/token ratio* of this sentence is 25/29 or 0.86. It is necessary to note that such a *ratio* decreases as the number of words increases. Therefore, the *ratio* given in this example cannot be compared with the *ratios* given for our samples in Table 1.

The version	*Type words*	*Token words*	*Type/token ratios*
Original	379	1152	0.32
Dubbed	397	1183	0.33
Subtitled	394	1134	0.34

Table.1: type/token ratios

It is clear from Table. 1 that the ratios of the three versions are somewhat almost similar. The subtitled version yielded: 0.34. This ratio is slightly higher than that of the dubbed version (0.33) and the original version (0.32). Comparing these ratios with the results of Chafe's comparison (1987) between "conversations, lectures, letters and academic papers" in terms of lexical variety, we see that our samples display more lexical variety. For instance, the ratio given for the original and dubbed versions are higher than those given for conversations and lectures in Chafe's study. This difference can be attributed to the fact that although the language used in both the original and dubbed versions is spoken, this language is not spontaneous. The conversations used in Chafe's study are taken from spontaneous spoken language. As a result, they show less lexical variety in comparison with written genres. By contrast, the conversations in our samples are not spontaneous though they share many features with spontaneous conversations.

THE USE OF VAGUE EXPRESSIONS

Inexplicit third person reference

The use of third person neuter pronouns "it", "this" and "that" is a feature of vague language. The specificity of these pronouns is that they do not have an explicit antecedent. They are frequent in spontaneous spoken language. This is because speakers cannot or do not take time to be explicit about what they are referring to. Table 2 shows the frequency of inexplicit third person pronouns in our data, in terms of the mean number of occurrences per million words.

The version	***The ratio***
The original version:	15625
The dubbed version:	16060, 60
The subtitled version:	14109, 34

Table 2: Inexplicit third person reference.

These statistics are represented in graphic form in fig.1.

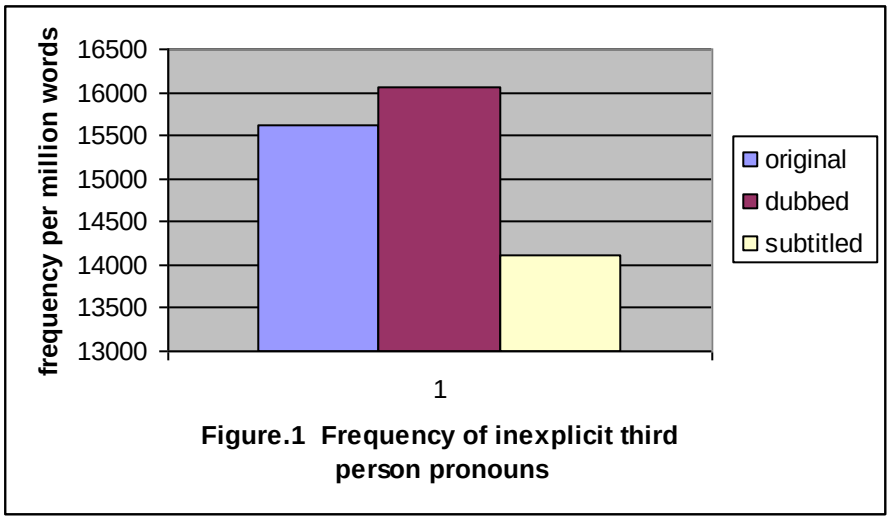

Figure.1 Frequency of inexplicit third person pronouns

It can be seen from table .2 and its graphic representation that the frequency of inexplicit third person pronouns in the original and dubbed versions is higher than that in the subtitled version. Chafe (1987) attributed the use of vague expressions, such as inexplicit third person pronouns, to the speaker's awareness of his limitation in choosing lexical items. In our case, we can attribute this higher frequency of inexplicit third person pronouns in the original and dubbed versions to the attempt to make the original language and dubbed language more authentic and spontaneous-like by including features which are characteristic of spontaneous spoken language. For illustration, we provide the following examples from our data:

1: ……. But I have tracked him here to los Angeles. He is with another company agent that I know as Gretchen.

……. the same agent who murdered Sara.

*……. **this** ends to day*

2: …….. who you buying it for?

A competitor

A foreign country

*Jason, you know I can't get into **that***

3: ……. The company will do anything to get it back.

*…….. Don't I know **it**?*

(« Prison Break »Original version)

4: …….. Alors ou est-elle?

***Ca** je te le dirai lorsqu'on sera dehors.*

(« Prison Break » Dubbed version)

5 : vous avez insisté pour qu'on le sorte de Sona.

*Il **le** fallait.*

(« Prison Break »Subtitled version)

What is specific about all the underlined pronouns in these examples is that they do not refer to an explicit noun phrase. They do not have an explicit antecedent. They refer, instead, to ideas talked about in the previous lines. For example, the pronoun "le" in "*je te le dirai*" refers to "saying where Sara is". This idea is not explicitly stated in a noun phrase. The pronoun refers to the answer of the question "ou est-elle?".

LEVEL OF LEXICAL ITEMS:

In English and French, we generally distinguish between three levels of language (trois niveaux de langue): colloquial (familier), neutral (courant) and formal (soutenu). The use of a particular level is dependent on the language use and the context in which the language is used. The language that is used in the original version is mainly colloquial. This contributes to the authenticity of this language and makes it similar to the spontaneous spoken language. In this case study, I will classify lexical items into two major classes: colloquial and literary. Table .3 shows the number of occurrences per million words of distinctly literary or distinctly colloquial vocabulary.

The version	Colloquial vocabulary	Literary vocabulary
Original	11284, 72	3472, 22
Dubbed	9298, 39	2535, 92
Subtitled	7054, 67	2645, 50

Table3: literary and colloquial vocabulary. Table .3 is graphically represented in Fig.2

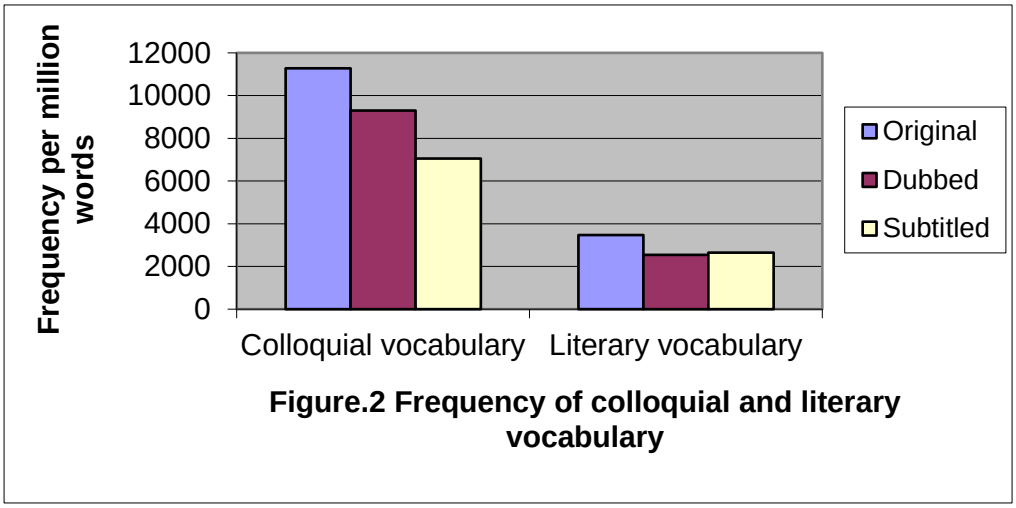

Figure.2 Frequency of colloquial and literary vocabulary

Comparing the frequencies of colloquial vocabulary in the three versions, it can be seen that the original and dubbed versions have higher ratios than the subtitled version. This result can be explained by the fact that the frequent use of colloquial lexical items is a feature of spoken language. By contrast, such a feature is less frequent in written language. However, if we contrast the use of colloquial vocabulary to that of literary vocabulary, we find that the ratios of colloquial lexical items are higher than those of literary lexical items in the subtitled version. This can be attributed to the subtitler's attempt to remain faithful to the register used in the original version. For the sake of illustration, we provide the following examples:

Original version	***Dubbed version***	***Subtitled version***
1*: "we get the hell out of there…"*	On se tire	On se case
2: "*you worry about getting your ass back in that seat in time….*"	Occupe-toi plutôt de ramener tes fesses dedans à temps….	Inquiétez vous plutôt de revenir vous asseoir ici à l'heure.
3 :Shut up	La ferme	La ferme

In the first example, the expression used in the original version is colloquial. The lexical item "*hell*" is used in its informal sense. This informality is retained in both the dubbed and subtitled versions through the use of the colloquial expressions (la langue familiere): "*on se tire*" and "*on se casse*". In the second example, the use of the colloquial lexical item "*ass*" reflects the extreme informality of the exchange. With its original lexical meaning (a part of the body), "*ass*" is usually a marker of informality. For instance, when speakers choose to use **"*ass*"** instead of **"*buttocks*",** they indicate that they hope to keep the tone of conversation informal and casual. The word "*ass"* is translated in the dubbed version into "*fesses",* which retains the informality expressed in the original version. The word "*ass"* is deleted in the subtitled sample and thus informality of the original version is not lost. The subtitler focuses on conveying the meaning rather than maintaining the feature of informality. This can be due to the fact that the lexical item "*ass"* is very informal and would not be acceptable in subtitling. In the third example, the lexical item "*shut up"* is translated in both versions into the colloquial (familier) "*la ferme".*

There is another feature of spoken language which is similar to the use of colloquial vocabulary. This feature is the use of repeats. Such a feature is rare in academic writings. Table 4 shows the number of occurrences of repeats per million words:

The original version:	6076, 38
The dubbed version:	5071, 85
The subtitled version:	1763, 66

Table 4: frequency of repeats per million words.

The graphic representation of Table.4

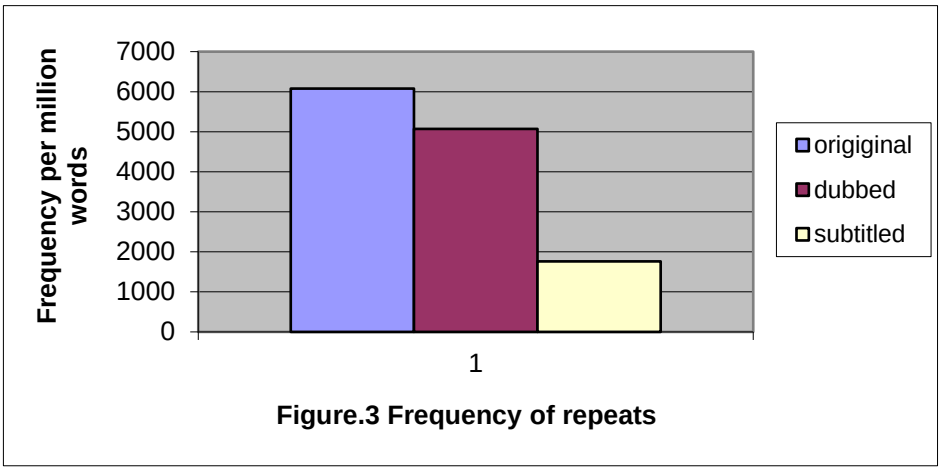

Figure.3 Frequency of repeats

It can be seen from Figure.3 that the phenomenon of repeats is almost equally present in the original and dubbed versions. Whereas its frequency in the subtitled version is considerably lower. This is due to the fact that repeats are a feature of spoken language and are not, generally, accepted in written language. Their frequent use in the original and dubbed version contributes to the authenticity of the language used in both versions. Let's have a look at the following examples from our data:

Original version	Dubbed version	Subtitled version
1: *"no, I'm, I'm safe"*	*Non ! c'est bon. Tout va bien*	*je vais bien*
2 : '' *thank god* ''	*T'en mieux, t'en mieux*	*Dieu merci*
3 : *''let's just….let's just start a new life''*	*On t'attend… on commence une nouvelle vie*	*Commonçons… Commonçons une nouvelle vie*

In the first example, the phenomenon of repeats present in the original version is retained neither in the dubbed nor in the subtitled version. In the second example, the feature of repeats used in the original version (*thank god, thank god*) is maintained in the dubbed version (*t'en meiux, t'en mieux*). However, the expression used does not express exactly the same meaning as the original expression. In the subtitled version, we notice the use of the exact equivalent of the original expression (*dieux merci)*. The repeat is not maintained because it is a feature of spoken language and the primary objective of subtitling is to convey the meaning expressed in the original version using the minimum of words.

INVOLVED VS. INFORMATIONAL FEATURES:

Biber classifies the features of Dimension.1 into two major classes: *positive and negative features.* The positive features characterize "involved" registers. In our case study, we measure the frequency of the following positive features: first- and second- person pronouns, private verbs, and present-tense. The negative features typify "informational" registers. Among these features, we count the frequency of the following: *prepositional phrases, normalisations and attributive adjectives.*

Positive features:

First- and second- person pronouns

Table. 5 includes in addition to first and second person pronouns, the third person personal pronouns in order to provide a better overall picture of personal pronoun use in our data.

The version	First-person pronouns	Second-person pronouns	Third-person pronouns
Original version	64236, 11	48611, 11	32986, 11
Dubbed version	71005, 91	49873, 20	29585, 79
Subtitled version	56437, 38	56437, 38	32627, 86

Table5: frequency of personal pronouns, listed in terms of the mean number of occurrences per million words.

The statistics in Table.5 are represented in graphic form in fig.4

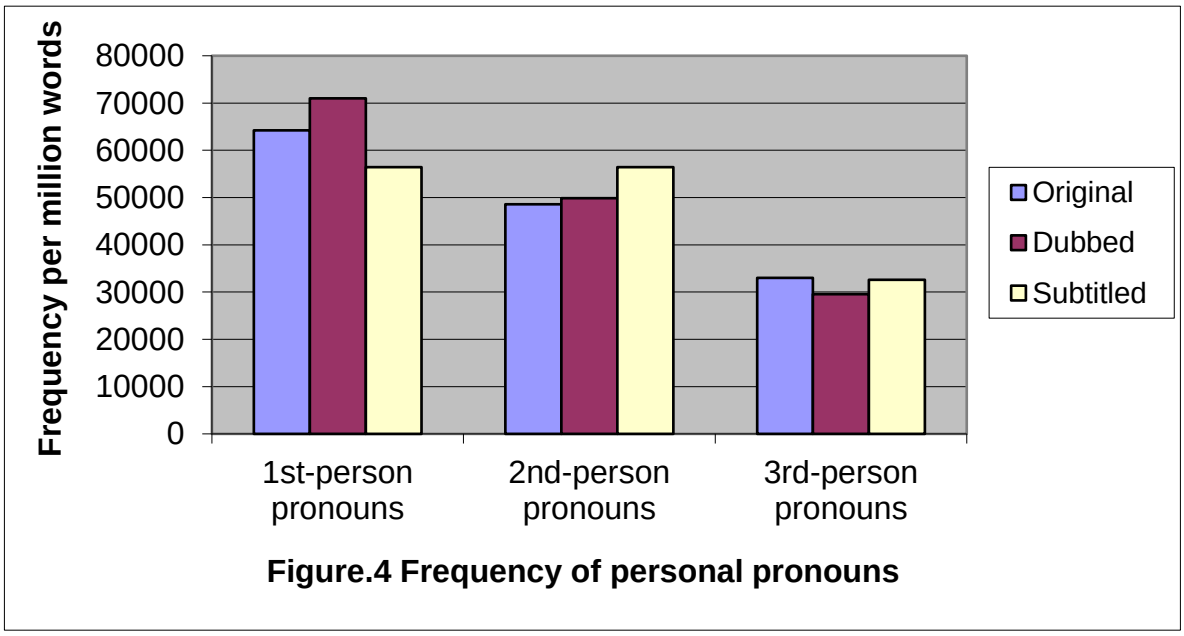

Figure.4 Frequency of personal pronouns

In the case of personal pronoun use, it will be interesting to analyse the proportion of first and second person pronouns to the number of third person pronouns within each sample. The ratio of first and second person pronouns (combined together) is higher in the dubbed version than the original and subtitled versions: 4: 1 (120882:29585.75). The ratio in the original and subtitled versions is nearly the same. In the original version, the ratio is 3.42: 1 (112847:32986), while in the subtitled version,

the ratio is 3.45:1 (112874:32627). This analysis shows that the three samples of our data have nearly similar frequencies of pronoun use. They all yielded higher frequencies of first and second person pronouns and lower frequencies of third person pronouns. This suggests that the registers used in the three versions can be considered as "*involved".*

The use of present tense:

The high frequency of present tense is another positive feature, which characterizes *involved registers.* Table. 6 shows the number of instances of present tense, listed again in terms of the mean number of occurrences per million words.

The original version	65104.16
The dubbed version	69319. 3
The subtitled version	70546.73

Table6: frequency of present tense per million words.

Table.6 is represented graphically in the following figure:

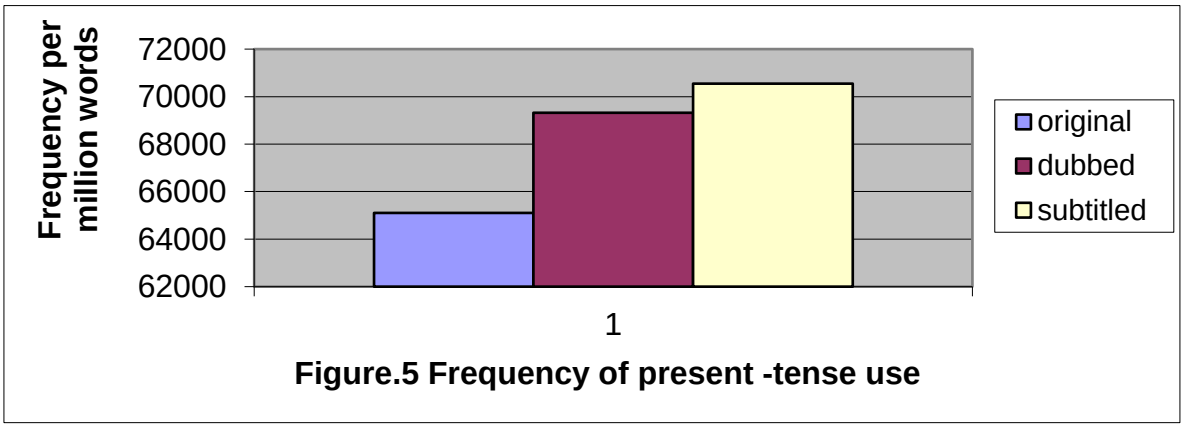

Figure.5 Frequency of present -tense use

We can see from figure.5 that the dubbed and subtitled versions have nearly similar frequencies of *present-tense* use. The ratio in the original version is slightly lower. This difference can be attributed to differences between English and French. In the English transcript, we have counted only the instances of *present-tense* and ignored the use of present continuous. In the French dubbed and subtitled versions, the present continuous is translated into *present tense* since the continuous aspect does not exist in French. This results in more instances of *present-tense* use in the dubbed and subtitled versions.

private verbs:

In addition to the frequent use of first- and second- person pronouns and present-tense, a similar important feature of "*involved"* registers is the frequent use of private verbs as opposed to public verbs. I have based my frequency counts of private verbs, in this case study, on **Douglas Biber's** definition: "*private verbs express intellectual states (e.g.: believe, know, and think) or non- observable intellectual verb (**e.g.:** discover).* In other words, they are used to express private feelings, thoughts and attitudes of which the speaker alone is aware (biber 1989). Table 7 shows the distribution of private verbs in our data.

Original version	37326, 38
Dubbed version	32121, 72
Subtitled version	36155, 20

Table 7: frequency of private verbs in terms of the mean number of occurrences per million words.

These statistics are graphically represented in fig.6.

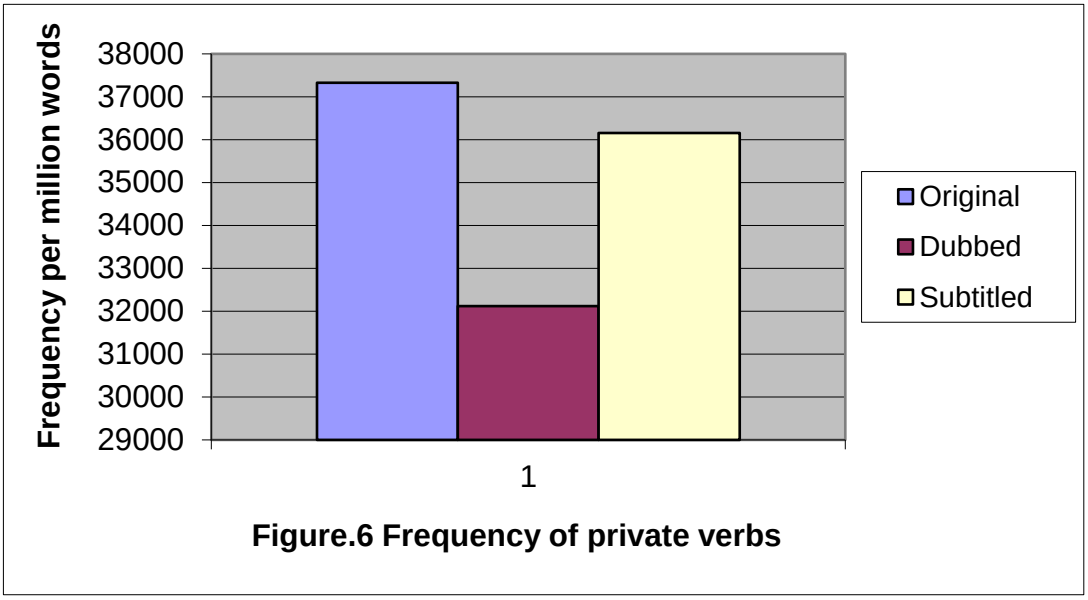

Figure.6 Frequency of private verbs

It is of interest here that the ratio of private verbs in the subtitled version is about the same as that in the original version. In the dubbed version, the ratio is lower than the ratios of the other two versions. This may be due to the constraints of dubbing. The dubber is sometimes obliged to delete some private verbs in his translation because of *lip-synchronization.* For a simple illustration, we provide the following example:

Original version	Dubbed version	Subtitled version
"Do you want to sell the card to me or not?…"	*"vous me vendez la carte oui ou non?..... »*	*Vous voulez me vendre la carte oui ou non ?*

In this example, the dubber opted for the deletion of the private verb, ***'want',*** because it does not match the lip-movement .Since Such a constraint is absent in subtitling, the subtitler opted for the translation of the original verb into the French equivalent private verb "*voulez*".

Negative features

Biber (1988) has identified many linguistic features which form the negative features of Dimension 1: *involved vs informational.* Those features are more frequent in written genres. In this study, I will consider only a few of the more common of them. I will count their frequency in our three samples and draw some conclusions. There are three such devices which are used much more frequently than others: *prepositional phrases, nominalizations and attributive adjectives.* The distribution of prepositional phrases in our data is shown in table 8, in terms of the mean number of occurrences per million words.

Original version	52083, 33
Dubbed version	61707, 52
Subtitled version	74074

Table8: the number of occurrences of prepositional phrases per million words.

We provide the graphic representation of table.8 in fig.7.

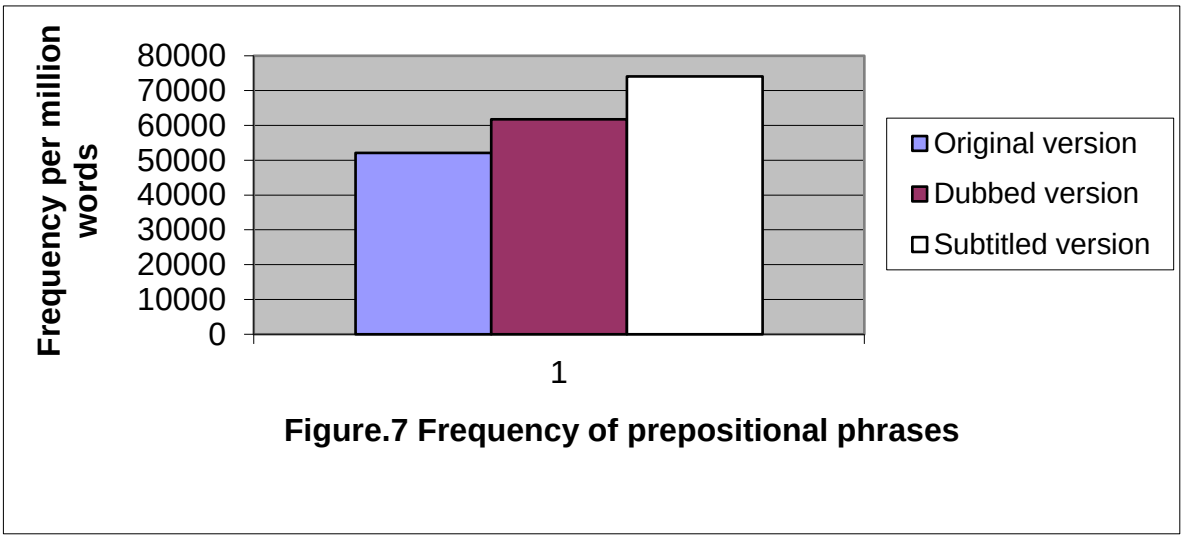

Figure.7 Frequency of prepositional phrases

Figure. 7 shows that the phenomenon of prepositional phrases is more frequent in the subtitled version than the other two versions. The ratio in the dubbed version is somewhat closer to that in the original version. This result has one explanation. This has to do with the nature of language used in each version. The language used in the original and dubbed versions is oral. The use of prepositional phrases in oral language is less frequent. By contrast, the language used in subtitling is written. This justifies the high frequency of prepositional phrases in the subtitled versions.

What is more, if we compare the number of occurrences of prepositional phrases in the original and dubbed versions to that in **Chafe's** study (1987), We see that they are somewhat similar (53 instances per thousand words in conversation). However, **Chafe** made it clear that the use of prepositional phrases in conversations is different from their use in academic writings. In conversations, prepositions usually form close constructions with verbs and sometimes can stand as utterances in themselves. The same thing can be said about the use of prepositional phrases in our data. The following examples are taken from our corpus. They illustrate for the use of prepositions in constructions that are close to verbs.

Original version	Dubbed version	Subtitled version
1:*"three weeks ago I was in Panamanian prison."*	*Il y a trois semaines, j'ai été en prison au panama.*	*Il y a trois semaines, j'ai été incarcéré dans une prison panaméenne.*
2 : *" involved in all levels of industry and government"*	*Infiltrée à tous les niveaux de l'économie et du pouvoir*	*Impliqué à tous les niveaux de l'économie et du gouvernement*

A second important feature which is mainly used to increase the length of what **Chafe** called "*idea units",* is the use of nominalizations, the formation of a noun from a verb. The distribution of nominalizations in our samples is shown in table. 9, listed in terms the mean number of occurrences per million words.

Original version	2604, 16
Dubbed version	3381,23
Subtitled version	3527

Table 9: The number of occurrences of nominalizations in the three versions

The graphic representation of table.9 is given in fig.8

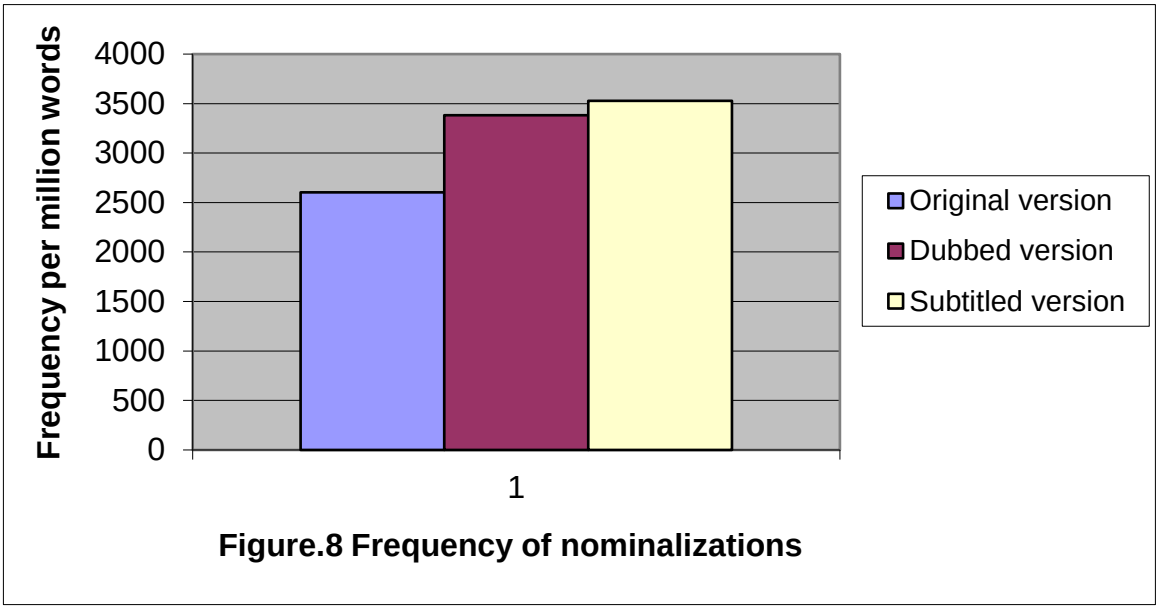

Figure.8 Frequency of nominalizations

The results in table .9 are somewhat similar to those in table.8. The ratio of nominalizations in the subtitled version is higher than that in the other two versions. However, the difference between the frequencies is not very significant. Moreover, comparing the ratios of nominalizations in this study with those in **Chafe's** study (1987), we see that the frequency in the subtitled version is somewhat closer to that in conversations (27 occurrences per thousand words). This has an explanation. The original language is oral and consists of oral conversations. Subtitles represent the translation of these conversations. Thus, although the language used in subtitling is written, features such as nominalizations are less frequent because these are also less frequent in the original language. The following examples illustrate the use of nominalizations in our data:

Original version	Dubbed version	Subtitled version
1:*"break one of their man James whistler out of that prison"*	***L'évasion*** *d'un de leurs hommes James Whistler*	*Faire sortir l'un de leurs hommes, James Whistler de cette prison*
2 : *"I came here seeking justice"*	*Je suis venu me faire justice*	*Je suis venu à la* ***recherche*** *de justice.*
3:*"planning the* ***invasion*** *of Normandy"*	*À planifier le* ***débarquement*** *de normandie*	*A orchestrer* ***l'invasion*** *de la normandie*

In the first example, the verb "*break out"* is nominalised in the dubbed version through the use of the noun "*l'évasion*" instead of using the verb "*evader*". Similarly, the present participle "*seeking",* in the second example is nominalized in the subtitled version using the noun "*la recherche"* instead of *"rechercher"*. In the third example, we have an instance of nominalization in the original version "*invasion".* This nominalization is retained in both the dubbed and subtitled versions.

A third device of similar sort is the use of attributive adjectives. These are used to describe or emphasize some characteristics of the noun they modify. In French, they are known as "*épithetes"* and they usually follow the noun they modify. Table.10 shows the number of occurrences of attributive adjectives in the three versions:

Original version	14756, 94
Dubbed version	11834, 31
Subtitled version	11463, 84

Table10. The mean number of occurrences of attributive adjectives per million words

The graphic representation of table.10 is given in fig.9.

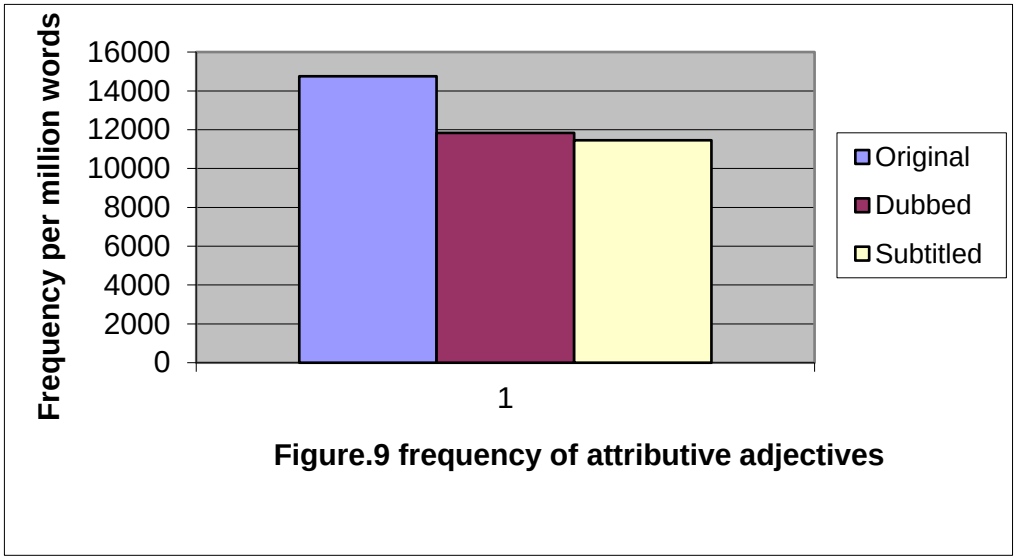

Figure.9 frequency of attributive adjectives

Table. 10 shows that the number of occurrences of attributive adjectives in the original version is higher than that in the other two versions. The ratios in the dubbed and subtitled versions are nearly the same. What is interesting here is the ratio yielded by the subtitled version. Although the language used in this version is written, attributive adjectives are less frequent. This may be explained into two ways. First, as we have already said the subtitled language is a translation of a spoken language. Therefore, it tends to include features of that language. Second, subtitling is required to be condensed and reductive. Therefore, the subtitler tries to avoid the excessive use of attributive adjectives as these will increase the length of subtitles.

Interpretation of the results

I have tried in this case study to measure the frequency of a number of oral and written linguistic features in the transcript of "*Prison Break", season 4, episode 1,* and its French dubbed and subtitled versions using **Douglas Biber's** Multidimensional Analysis (1988). First, in order to measure the variety of lexical items in our data, I used the mechanical measure *type/token ratio.* The results are shown in table.1. Depending on these results, we can say that the language used in the subtitled sample is more varied than that used in the original and dubbed samples. For the *type/token ratio* yielded by the subtitled version (0.34) is higher than that of the original and dubbed versions (the original has 0.32, while the dubbed version has 0.33). However, the difference between the ratios is not significant. The ratios are nearly the same in the three versions. I can attribute the similarity between the dubbed and the subtitled samples in terms of *type/token ratios* to the fact that both samples represent the same spoken source language. Therefore, they tend to share many linguistic features. The same thing can be said about the use of inexplicit third person pronouns. The results in table. 2 suggest that inexplicit third person pronouns are almost equally frequent in the three samples.

Original version	Dubbed version	Subtitled version
*1: **This** ends today*	*L'histoire prend fin aujourd'hui*	***ça** finit aujourd'hui*
*2: Don't I know **it**?*	*Oh, oui vraiment*	*Je ne **le** sais pas trop*

These examples illustrate the use of inexplicit third person pronouns in our data. In the first example the pronoun "*this"* does not have any explicit antecedent. It refers, instead, to a story that is told in the previous lines. This pronoun is transformed in the dubbed version into "*histoire"* which can be considered as an interpretation of the pronoun. As a result, the sense of vagueness created by this pronoun in the original expression is lost in the dubbed version. By contrast, this vagueness is retained in the subtitled version through the use of the French equivalent pronoun "*ça".*

In measuring the level of the lexical items in the three samples, I chose to count the frequency of literary and colloquial vocabulary. As it is shown in the table.3, colloquial lexical items are more frequent in the original and dubbed versions and slightly less frequent in the subtitled version. The ratios of the colloquial vocabulary are much higher than those of literary vocabulary in the three samples. Of interest here is the fact that the subtitled sample has the lowest ratio of both colloquial and literary vocabulary. This suggests that the subtitler prefers the use of neutral lexical items in the translation of original informal ones.

Original version	Dubbed version	Subtitled version
*"You worry about getting your **ass** back in that seat in time"*	"*occupe-toi de ramener tes **fesses** dedans à temps"*	*« Inquiétez vous plutôt de revenir **vous asseoir** ici à l'heure. »*

This example illustrates for the translation of a colloquial lexical item (*ass).* The use of this informal word reflects the informality of the original expression. This sense of informality is retained in the dubbed version through the use of the French equivalent of "*ass"*; *"fesses".* In the subtitled version, the same meaning is conveyed, but the sense of informality is lost. The subtitler opted for neutral lexical items *"revenir"; "vous assoir"*.

Another feature which is similar to the use of colloquial vocabulary is the use of repeats. We have seen in table. 4 that such a feature is more frequent in the original and dubbed samples. In the subtitled sample, the ratio of repeats is considerably lower. We can interpret this result by saying that the primary concern of the subtitler is to convey the meaning with a minimum of words. Thus, repeats and unnecessary expressions are likely to be deleted.

The use of such features suggests that we can make a distinction between two registers. One is referred to as "i*nvolved",* and the other as *"informational".* Each of these registers is characterized by specific linguistic features. Involved registers have high frequencies of positive features. Among these features we chose to count the frequency of the following: first and second person pronouns, present tense and private verbs. As for "*informational registers",* they are characterized by high frequencies of negative features. Three major devices of these features were studied in this study **e.g.:** prepositional phrases, nominalizations, and attributive adjectives. The following table presents the total number of both positive and negative features in the three samples of our data, listed in terms of the mean number per million words.

The versions	*Positive features:* *-**1st** and **2nd** person pronouns* *-present tense* *-Private verbs*	*negative features:* *-prepositional phrases* *-nominalization* *-attributive adjectives.*
Original	215277, 76	69444, 43
Dubbed	222316, 13	76923, 06
Subtitled	219576, 69	89064, 84

Table11: the total number of positive and negative features in the three versions

It is clear from the table above that there is a huge difference between the ratios of positive and negative features in the three versions. The frequencies of negative features are considerably lower when compared to those of positive features. In other words, the positive features are more frequent than the negative features within each version. Thus, the register used in each version is more *"involved"* and less *"informational".* What is more, although the language used in the subtitled sample is written, it has a very high ratio of positive features. Since positive and negative features are in complementary distribution,

the frequency of negative features is significantly lower. Thus, we can say that the oral features of the original language are transferred into the subtitled language despite the fact that this language is written.

After knowing that the use of *"involved"* features is more frequent than that of "*informational"* features within each version, it will be interesting to determine which version is more involved and which one is more informational. If we compare between the dubbed and subtitled versions in terms of the use of positive features, we see that the dubbed version has the higher frequency. As a result, the language used in this version is more "*involved"* than that used in the subtitled version. By contrast, the number of occurrences of negative features in the subtitled version is higher than that in the dubbed version. Thus, the register used in the subtitled language is more "*informational"* than that used in the dubbed language.

In summary, depending on the analysis of the data, a number of conclusions can be drawn:

1. Positive features are more frequent in the three samples, as opposed to negative features which are considerably less frequent. This proves Biber's concept that positive and negative features are in complementary distribution. That is, they cannot be used in a register with equal frequencies. The high frequencies of the first group (positive features) must lead to the low frequencies of the second group (negative features) and vice versa.
2. The register used within each sample of our corpus is highly involved (interactive). The features characterizing this register such as first- and second-person pronouns, private verbs and present-tense are highly frequent in the three versions.
3. The original and dubbed samples share many positive and negative features. This is due to the fact that the medium used in both samples is spoken language. As a result, they tend to favour oral linguistic features.
4. The subtitled language resembles the original and dubbed languages in that it is more involved. Although the medium used in the sample is written language, the subtitler tries to remain faithful to the register used in the original sample by including features that characterize this register. However, when comparing the degree of "*informationality"* in the three versions, we can conclude that the subtitled version is more "informational" than the other two versions. Thus, we can say that the fact that the medium used in the subtitled version is written affects to some extent the choices of the subtitler who tries to include some features that characterize this medium ("informational"features). Still, these features remain less frequent when compared to "involved" features.

Conclusion

I have attempted in this research to study the use of oral and written linguistic features in audio-visual translation, specifically in dubbing and subtitling. To achieve this purpose, I have investigated those issues by conducting a case study. I chose as a corpus for my study the transcript of "*Prison Break", season 4, episode 1,* and its French dubbed and subtitled versions. I have studied the use of oral and written linguistic features in the three versions. I counted the number of occurrences of a number of linguistic features. In doing so, I adopted **Biber's** Dimension 1: *involved vs. informational* production. According to this dimension, we distinguish between positive and negative features. The high frequency of the first group characterizes involved registers. Whereas, the high frequency of the second group typifies informational registers. In this case study, it can be concluded that positive features, namely first- and second person pronouns, private verbs and present -tense are more frequent than the negative features in the three samples. Thus, the register used in the three samples of our data is highly involved. But when comparing between the three versions, we see that the dubbed language is more involved than the subtitled language. Similarly, when comparing the degree of "informationality" in the three samples, we can see that the subtitled language is more informational than the original and dubbed languages.

Finally, I have to say that the findings of the present study are limited to our chosen corpus and should not be generalized to other corpora. However, I think this study strongly indicates that audio-visual translation has much potential to provide us with important data for linguistic analysis. I hope that this case study will encourage other researchers to devote more attention to this virtually unexplored, yet exciting research area.

References

[1] Andrew T. (2006). Media Talk: Spoken Discourse on TV and Radio. Edinburgh: Edinburgh University Press Ltd.

[2] Biber, D., Connor, U., & Upton, T. (2007). Discourse on the Move. *Using corpus analysis to describe discourse structure*. Amsterdam/Philadelphia: John Benjamins Publishing Company

[3] Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

[4] Biber, D. (2004). Conversation Text Types : A multi-dimensional analysis. Northern Arizona : Jadt (2004) : 7emes Journées Internationales d'analyse statistique des données textuelles.

[5] Brown, G., Brown, G. D., Brown, G. R., Gillian, B., & Yule, G. (1983). *Discourse analysis*. Cambridge university press.

[6] Chafe, Wallace & Jane, Danielewics (1987). Properties of spoken and Written Language. In Horowitz R., Samuels F.J., Comprehending Oral and Written Language. New York: Academic Press.

[7] Chafe, Wallace (1982). Integration and involvement in speaking, Writing and oral literatures. In Tannen D., Spoken and Written Language: Exploring Orality and Literacy. Norwood: Albex.

[8] Chafe, Wallace (1994) Discourse, consciousness and time: the flow and displacement of conscious experience in speaking and writing. Chicago: the University of Chicago Press.

[9] David, B. & Kristin, Th. (1979). Film Art: an Introduction. New York: McGraw Hill.

[10] Deborah T. (1982). The oral/Literal Continuum in Discourse. In Tannen, Deborah: Spoken and Written Language: Exploring Orality and Lteracy. Norwood: Albex.

[11] Deborah T. (1988). The commingling of Orality and Literacy in Giving a Paper at a Scholarly conference. American Speech 63.

[12] Delia C., Christine H. & Bucaria C. (2008). Between Text and Image, Updating research in screen Translation. Amsterdam/Philadelphia: John Benjamin is Publishing Company.

[13] Frederic Ch. (2004). Film Studies and Translation Studies: Two disciplines at Stake in Audio-visual Translation. Meta: Translators' Journal, Vol.49 nº1.

[14] Jenney, Mattson (2006). Linguistic Variation in Subtitling: The Subtitling of Swearwords and Discourse Markers on Public Television, Commercial Television and DVD. In Audio-visual Translation Scenarios: Conference Proceedings. Muttra.

[15] Jim, M. & Reigina W. (1998). Spontaneous Spoken Language: Syntax and Discourse. New York: Oxford Clarendon Press.

[16] Taylor, C. J. (2004). The language of film: corpora and statistics in the search for authenticity. Notting Hill (1998), a case study. *miscelánea: a journal of english and american studies*, (30), 71-86.

[17] Jorge D. C. (2009). *New Trends in Audio-visual Translation (Topics in Translation)*. London: Multilingual Matters.

[18] Jorge D.D., Guinila M.A (2009). Audio-visual Translation: Language Transfer on Screen. Hampshire: Palgrave Macmillan, 2009.

[19] Mera M (1991). Read my lips: Re-evaluating subtitling and dubbing in Europe. Letters and Links 6.

[20] Mona B. (2005). *Routledge Encyclopaedia of translation studies*. London and New York: Routledge.

[21] Orero P. (2004). *Topics in audio-visual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

[22] Per L. (2005). The Written Language Bias in Linguistics: Its Nature, Origins and Transformations. London and New York: Routledge.

[23] Pettit, Z. (2005). Translating register, style and tone in dubbing and subtitling. *Journal of Specialised Translation*, *4*(4), 49-65.

[24] Pavesi, M. (2008). Spoken language in film dubbing Target language norms, interference. *Between text and image: Updating research in screen translation*, *78*, 79.

[25] Quaglio P. (2009). *Television Dialogue: The Sitcom Friends vs. Natural Conversation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.