



A quantitative analysis of racist epithets referring to Italians and their translations in movie subtitles: The case of *wop*, *eyetie* and *goombah*

Serena Ghiselli

University of Bologna

Forlì, Italy

serena.ghiselli3@unibo.it

<https://orcid.org/0000-0002-1252-4018> 

Abstract: Anglophone cultures have had extensive contact with other ethnic groups over the last 200 years and almost always from a position of dominance (Filmer, 2012). In English there are racial slurs for practically every race or ethnic group and swear words are commonly used in English TV dialogues. Translating swear words in movie subtitles pose a translation challenge, which has been studied by scholars such as Soler Pardo (2015), Ávila-Cabrera (2016), Beseghi (2016) and Díaz-Pérez (2020). This paper analyses the translation from English into Italian of *wop*, *eyetie* and *goombah*, three racist epithets used to refer to Italians or people of Italian origin. The data analyzed are taken from the English-Italian parallel corpus of *OpenSubtitles* (Lison & Tiedemann, 2016), a collection of parallel corpora made up of translated movie subtitles. English subtitles from a variety of movies including these racist epithets are retrieved, together with the corresponding Italian translations, using parallel concordance. The renditions are analyzed with the aim of identifying general trends in the use and in the translation of these racial slurs. In addition, renditions are divided according to moderating variables such as the languages spoken in the film, the film release date, the countries the film was shot in and the genres of the film, to test whether these variables have a significant impact on translation choices.

Keywords: racial slurs; translation; subtitles; English; Italian.

1. Introduction

Racial slurs are offensive and derogatory words or phrases that target individuals or groups based on their race, ethnicity, or nationality. These racist epithets are based on xenophobic attitudes and prejudice against people who are not part of the dominant ethnic group in a society. The



stereotypes are rooted in traditional beliefs and biased attitudes, as they consistently portray negative traits, often emphasizing behaviors like laziness, uncleanness, inefficiency, foolishness, cruelty, cowardice, aggression, alcoholism, promiscuity, and deviance (Hughes, 2006). The identity of others in insults is often depicted in a stereotypical way by *blasons populaires*, which can be defined as badges given to a group by outsiders, which imply that all the members of the targeted group are homogeneous and share the same failings (Hughes, 2006).

Over the past two centuries, English-speaking cultures have engaged in interactions with various ethnic groups, typically from a position of power and influence (Filmer, 2012) and language has played a major role in the reinforcement of ethnic stereotypes. Racial slurs express verbally a racist attitude towards a certain social group, and Memmi (1999) considers racism as the general assessment of actual or imaginary distinctions that favor the accuser while harming their victim. The accuser uses this negative judgment to justify their privileges or aggressive behavior. He proposes two definitions of racism: the narrow definition only refers to a discriminatory evaluation that includes both real and fictitious biological differences; the broad definition is what he describes with the neologism *heterophobia*, that is the “many configurations of fear, hate, and aggressiveness that, directed against an other, attempt to justify themselves through different psychological, cultural, social or metaphysical means, of which racism in its biological sense is only one instance” (Memmi, 1999, p. 118). According to the social identity approach (Hogg & Abrams, 1998), dividing society into groups is a way to reduce its complexity, and socially acquired patterns of perception tend to exhibit a preference for one's own ingroup while displaying derogatory attitudes towards outgroups (Reisigl & Wodak, 2001).

2. Racist epithets in English: the case of people of Italian origin

Throughout history, English perceptions of Italy and its inhabitants have been ambivalent, showcasing a blend of admiration and aversion shaped by cultural affiliations and religious differences (Hughes, 2006). Allen (1983) studied the numerous nicknames that Italian immigrants were given in the United States, where they were seen as aliens and outsiders. Among these nicknames there are *wop* and *eyetie*. Folk etymology, that is a plausible but inaccurate explanation of the origin of a word, claims that the racial slur *wop* is also an acronym, derived from an immigration category WithOut Passport or WithOut Papers, alternatively Working on Pavement (Hughes, 2006). These derivations are based on a negative stereotype of Italians as illegal immigrants or low-status laborers. However, as Allen (1983) underlined, these etymologies are not coherent with the context in which *wop* first appeared:

First, all immigrants without documentation would have been nicknamed the same, but Italians were the only immigrant group in the 1890s and later who were called wops. Secondly, the nickname emerged in American slang [...] before acronyms came into wide use in government bureaucracies (Allen, 1983, p. 119).

Wop dates from the 1890s and a more plausible etymology for this word is that it comes from Italian *guappo*, a word that in Neapolitan and Sicilian means “a dude, a swell, or a bold showy



ruffian” (Hughes, 2006, p. 258). Initially, after the peak of Italian immigration in the late 1800s, *guappo* would frequently be heard among Italian immigrants while teasing, and it was not perceived as derogatory (Delgado & Stefancic, 2004). By contrast, the first recorded use of *wop*, then spelled as *wap*, has a negative connotation and is explicitly associated with organized crime, as reported by Arthur Train (1912), a former Manhattan assistant district attorney, in *Courts, Criminals, and the Camorra*:

Curiously enough, there is a society of criminal young men in New York City who are almost the exact counterpart of the Apaches of Paris. They are known by the euphonious name, “W[o]ps” and “Jacks.” These are young Italian-Americans who allow themselves to be supported by one or two women, almost never of their race. These pimps affect a peculiar cut of hair, and dress with half-turned-up velvet collar, not unlike the old-time Camorrist [the Camorra being the Neapolitan version of the Mafia], and have manners and customs of their own. They frequent the lowest order of dance-halls, and are easily known by their picturesque styles of dancing, of which the most popular is the “Nigger.” They form one variety of the many “gangs” that infest the city, are quick to flash a knife as the Apaches, and, as a cult by themselves, form an interesting sociological study (Train, 1912, p. 232).

Among the numerous variations of the term *Italian*, *Eyetalian* was first documented in 1840. However, *eyetie* emerged much later, specifically after World War I. *Eyetie*, which is generally considered less offensive than *wop*, is an ironic diminutive with a phonetic similarity with the neutral term referring to that ethnic group, Italian, a feature which is not uncommon in xenophobic nicknames (Hughes, 2006).

Goombah is defined as a friend, a trusted associate, especially of organized crime; in this sense it is a synonym of *mafioso* (Chapman, 1986). It is not a derogatory term if used among Italian-American men to refer to a close friend or associate, but it becomes offensive when employed to demean or patronize someone of Italian-American descent (Merriam-Webster, n.d.). The origin of *goombah* can be traced back to the southern Italian words *cumpa* or *cumpari* and the standard Italian equivalent, *compare*, all meaning *godfather* (Treccani, n.d.).

3. Translating derogatory language in movie subtitles

Insults give information about social relations, which are crucial to understand where a person comes from (Greenall, 2011). In Italian there is a reduced variety of racial slurs compared to the British, American and Australian varieties of English (Filmer, 2012). This is probably due to the fact that multiculturalism is a more recent phenomenon in the Italian society compared to those of the Anglophone cultures, even if in recent decades also Italy has experienced an increase in immigration from several areas, such as China, North Africa and Eastern Europe (Filmer, 2012).

Even if the variety of derogatory expressions available differs from language to language, the use of offensive ethnic slurs is a practice that exists across various cultures. When translators have to convey their meaning in a different language and culture, they face two major challenges. Firstly, such ethnic slurs or derogatory expressions might be misunderstood or not comprehended at all by the audience in the target language. Secondly, these offensive terms become unacceptable to others, especially to the individuals referred or alluded to (Chamizo Domínguez, 2018). Chamizo



Domínguez (2018), after analyzing examples from TV series, comics and books, found out that in the attempt to adapt the source text to the target culture, the slur is removed in the target language either consciously or unconsciously.

As Ávila-Cabrera (2016) highlights, the translation of insults in subtitling has not been studied extensively so far. The existing contributions focus on specific films (Filmer, 2012), TV series (Beseghi, 2016) or on a specific film director (Soler Pardo, 2015), and there are few studies which use corpora of more varied audiovisual products to approach the issue of translating derogatory expressions in general and racist epithets in particular. The 2011 version of the corpus *OpenSubtitles* (Tiedemann, 2009) was used to study the translation of taboo words from English into Spanish, in particular *fuck* and *shit* (Torres Cuenca, 2016). The analysis of the translations showed that in 43% of cases the word was either omitted or softened. Díaz-Pérez (2020) used the *Veiga Corpus* (Sotelo, 2015) to analyze the translation of the same swear words, *fuck* and *shit*, from English into Galician. This second study highlighted once again a sanitizing tendency in translation, because in 49% of the cases the taboo word was omitted or neutralized and in 12% of the cases it was softened. Both authors argue that these results may be due to the need to shorten subtitles because of technical constraints, but also by the fact that taboo words can appear much ruder in written than in spoken form.

4. Research questions and aims

The close relation between language and society in the use of racial slurs calls for a reflection when they have to be translated into a different language and culture. This paper focuses on the use and on the corresponding translation into Italian of three English racist epithets for people of Italian origin (*wop*, *eyetie* and *goombah*) in the movie subtitles that form the parallel corpus *OpenSubtitles 2018 parallel - English*. The aim of the study is seeing whether the trend to censor offensive words observed by Chamizo Domínguez (2018) applies to *wop*, *eyetie* and *goombah* when subtitling into Italian. Furthermore, the study will try and establish whether translation choices are influenced by the type of ethnic slur used or by features of the context of production of the movie accessible to the researcher through the corpus metadata, namely film release date, countries the film was shot in, genres of the film and the languages spoken. For the film release date, it was hypothesized that newer films could potentially contain a higher number of racist epithets because of a reduced influence of censorship both in the original and in the translated subtitles. The fact that the film was shot in more than one country or in more than one language could imply that it deals with an intercultural topic. It is interesting to see whether the shooting of a film in multiple countries and/or languages corresponds to a higher or, on the contrary, to a reduced number of translations that keep the same tone and pragmatic function of the epithet. Eventually, the film genre was also analyzed to give a complete overview on the topic using the metadata available.



5. Materials and methods

The racist epithets and their translations were extracted from *OpenSubtitles 2018 parallel - English* (about 1.2 billion tokens), a collection of 60 sentence-aligned parallel corpora composed of movie subtitles¹. *OpenSubtitles*² constitutes a growing collection of authentic English subtitles along with their versions in different languages, making it likely the most extensive assortment of subtitles among freely available parallel linguistic corpora. The corpus is freely accessible through *Sketch Engine* (Kilgariff et al., 2014), an online corpus tool to create and search text corpora. This corpus was chosen because it is a large collection of subtitles that can be queried easily and for free. A negative aspect is that everyone can upload subtitles on *OpenSubtitles*, so there is no quality check on their content. In addition, the parallel concordance of *Sketch Engine* only shows limited context for the element searched and does not provide access to the video for which the subtitle was created. This makes a thorough qualitative evaluation of the subtitles and their translations impossible. Therefore, the analyses in this paper are entirely quantitative. The corpus includes meta-data about the audiovisual product for which the subtitle was created.

In the corpus, the films are accompanied by a set of meta data, categories, which were taken into consideration in the comparison between source and target text. The metadata considered are the following: the languages spoken in the film, the film release date, the countries the film was shot in and the genres of the film. The film title was also collected and used to find metadata by an Internet search when they were missing in the corpus.

Parallel concordance allowed the researcher to extract the English segments containing the racist epithets along with their corresponding translations into Italian. Three queries were used, one for each epithet: *wop*, *eyetie* and *goombah*; the Advanced query options were set to 'lemma' and 'case-insensitive' to retrieve all forms of words in the results. The parallel concordance was looked up in the aligned corpus of Italian subtitles. The queries found the following results: 197 hits for *wop*, 11 hits for *eyetie* and 49 hits for *goombah*. The results retrieved automatically were downloaded and manually analyzed.

Source and target texts were manually compared by the author to assign the translations to different categories according to the rendition of the racist epithets. The classification of translation solutions was taken from Díaz-Pérez (2020, p. 404), who categorized the translation of swear words as follows: *pragmatic correspondence*, *softening*, *de-swearing* and *omission*. *Pragmatic correspondence* indicates that the slur has been translated in the target text keeping the same tone and pragmatic function, ex. "EN: <s> Is that wop or nigger? </s>; IT: <s> Mangiaspaghetti o negro? </s>". *Softening* refers to a translation choice where the target text has a softer or milder tone, ex. "EN: "<s> -Never marry a Wop . </s>; IT: <s> Mai sposare un maccherone! </s>". With *de-swearing*, the meaning of the derogatory term is conveyed to the target text but using a non-offensive expression, ex. "EN: <s> No point in talking with the Eyeties. </s>; IT: <s> È inutile parlare con gli italiani. </s>". Eventually, *omission* means that the derogatory term does not appear in the target text, ex. "EN: <s> How much can a goombah like you have? </s>; IT: <s> Quanti soldi hai? </s>". The first two categories (*pragmatic*

¹ Cf. <http://opensubtitles.org>.

² This is part of a wider project called *Open Parallel Corpus (OPUS)* (Tiedemann, 2012), whose goal is to be a publicly available resource of parallel corpora that can be accessed for free.

correspondence and *softening*) can be defined as renditions that are more loyal to the source text, whereas the use of the other two categories (*de-swearing* and *omission*) indicates that the translator is removing the racist content conveyed by the slur.

While categorizing the translations, some hits were deleted because they could not be analyzed. For the epithet *wop*, 34 results were deleted because the term was left in English in the Italian text (ex. EN: *<s> -Italians– They got guinea, wop, dago. </s>*; IT: *<s> - Gli Italiani... </s><s> loro hanno Guinea, Wop, Dago. </s>*), or was not a racist epithet but part of an onomatopoeic expression (ex. *<s> Shoo wop doo wop shoo wop doo wop shoo wop doo wop shoo wop doo wop shoo wop doo wop shoo wop doo wop shoo wop doo wop ah </s>*). For *eyetie*, one hit was deleted because the term had not been translated: “EN: *<s> Eyetie, where in the hell is the van? </s>*; IT: *<s> Eye-tie, dove diavolo è il furgone? </s>*”. Seven hits of *goombah* were not analyzed because the term was left untranslated (ex. EN: *<s> is goombah's still open? </s>*; IT: *<s> Goombah è ancora aperto? </s>*). Although leaving the English term is a possible translation strategy in subtitling (e.g. Gottlieb, 1992), the three terms under investigation are totally unintelligible to an Italian audience and can be easily confused with proper names or nicknames. It is more likely that they were left due to a mistake rather than a deliberate choice on part of the translator. For this reason they are not considered in the analysis.

As regards the other variables considered in the analysis – the languages spoken in the film, the film release date, the countries the film was shot in and the genres of the film – they were treated as follows.

All the movies were in English, but some of them had one or more additional languages in the original dialogues. This variable was coded as the dichotomous variable *additional language*, which could be either *yes* or *no*. This variable was coded to see whether in a multilingual film the approach to the translation of racist epithets was different compared to a monolingual film.

Every corpus hit had a single option for the variable *date*, indicating the release date. When preparing the data for the analysis, the problem to tackle was that there were 43 different dates, so a simplification was needed. Too many categories meant that there would have been a tiny number of hits per date and, with a small dataset, statistical analysis would have been impossible. Moreover, it is conceptually unlikely that the use of racist epithets would differ in movies released in two consecutive years; if a difference exists it is logic to see it when comparing audiovisual products released in periods more distant in time. The dates of the movies were therefore divided into six ranges: before 1970, 1970-1979, 1980-1989, 1990-1999, 2000-2009 and 2010-2017 (there were no hits of a later date).

The variable *country* was given a label including all the places where the film was shot. For example, a film shot only in the United States was given the label *USA*, whereas a movie shot both in the United States and in the United Kingdom was labelled as *USAUK*. The single countries and the combination of two or three countries resulted overall in 16 categories, which remained a manageable number for the analysis.

Differently from *country*, the variable *genre* was rather challenging to handle in the analysis because it is a categorical variable that has multiple options per hits in most cases. It was not possible to use combined labels because there were too many possible combinations. With this approach there would have been 55 genre categories for 257 hits originally retrieved, which would have made

group comparison not feasible. As mentioned above for *date*, dividing up a small dataset into too many categories reduces the number of hits per group too much to make any comparison. At the same time, including only the first genre recorded in the corpus per hit was considered to be an excessive simplification, which omitted many examples of use of a certain slur matched with a specific genre. The only combination of two genre categories into a single one was done for the genres *music* and *musical*, which were both included under *musical*.

To compare translations without losing genre categories, a separate dataset was created. In this dataset, used only for the comparison of translation categories divided by genre, every hit having multiple options was repeated so that every data point had only one genre. In order to keep track of the fact that that specific string was originally a single hit, the variable *ID* was added to identify every original hit with a single label. Instead, the analysis for all the other variables (date, country and additional language) was carried out using the original dataset, cleaned from irrelevant hits.

First of all, the use of ethnic slurs in the corpus was described commenting on Table 1, which reports the distribution of the hits of *wop*, *eyetie* and *goombah* in different subcorpora. Then a descriptive analysis of the distribution of translation strategies by ethnic slur was carried out with pie charts and pivot tables using Microsoft Excel. Balloon plots were then created using the integrated development environment *R Studio 4.2.1* (RStudio, 2022), specifically the packages *ggplot2* and *ggpubr*, to represent the distribution of the translation categories in the different groups of the independent variables (the film release date, the countries the film was shot in, the genre of the film and whether there were additional languages spoken in the movie). A balloon plot is a graph that enhances the visualization of numerical information in contingency tables using colored circles whose sizes correspond to the magnitudes of the respective table elements. This representation effectively emphasizes key data characteristics while retaining the details provided by the numeric values (Jain & Warnes, 2006).

Finally, a statistical analysis was performed using *R Studio 4.2.1* (RStudio, 2022), specifically the packages *stats* and *effectsize*. The goal of the analysis was to explore whether the translation categories – the dependent variable – significantly differ depending on the racist epithet translated (*wop*, *eyetie*, *goombah*) and when the data are grouped by various independent variables. Separate group comparisons were carried out between the dependent variable and every independent variable.

All the variables are categorical and the comparisons were carried out using Pearson's Chi-squared test with Monte Carlo simulation (Hope, 1968). The chi-squared (χ^2) test is a statistical test used to determine whether there is a significant association between categorical variables (The University of Utah, 2023) and it was chosen because all the variables analyzed are categorical and the aim of this paper is to compare them to check whether the translation approach differs between the groups.

Before comparing the variables the independence (the observations in each category should be independent of each other), random sampling (the sample is representative of the population of interest), expected cell frequency (it is recommended that no more than 20% of the cells of the contingency table have expected frequencies below 5, and no cells have expected frequencies below 1) and sample size (large enough for the statistical analysis) assumptions were checked (Howitt & Cramer, 2020). In the dataset, with the exception of *additional language*, which is a dichotomous

variable, all the variables have more than two categories and the assumption of expected cell frequency in the contingency table (a table in a matrix format displaying the frequency distribution of the variables) was violated. Therefore, an adjustment was needed to reach statistically reliable results. Pearson's Chi-squared test with Monte Carlo simulation (or simulated *p*-value) based on 2000 replicates was performed. Monte Carlo simulation involves generating a large number of simulated contingency tables that satisfy the null hypothesis of independence between the variables and then comparing the observed test statistic with the simulated distribution to obtain a *p*-value. Monte Carlo simulation can handle small expected cell frequencies and contingency tables with more than two categories (Foley, 2023).

Eventually, the analysis considered not only statistical significance (*p* < 0.05) but also effect size. Effect size is a numerical representation indicating the magnitude of a phenomenon's significance, employed to investigate a matter of interest (Kelley & Preacher, 2012). The effect size calculated for the chi-squared test was Cramer's *V*, which measures the strength of association between two categorical variables (Bobbitt, 2023). It ranges from 0 (no association) to 1 (perfect association). Cramer's *V* is a commonly used effect size measure for chi-squared tests and is defined as the square root of the chi-squared statistic divided by the total number of observations, adjusted for the table dimensions. A general guideline to interpret Cramer's *V* is that an effect ≤ 0.1 can be considered small, an effect between 0.2 and 0.5 is moderate and an effect ≥ 0.5 is large.

6. Results

The first step of the analysis was looking at the number of slurs retrieved from the corpus. After excluding irrelevant hits (see §Materials and methods) the overall number of epithets to be analyzed was 215, divided as follows: 163 hits of *wop* (75%), 10 hits of *eyetie* (5%) and 42 hits of *goombah* (19%). From the data collected, it results that *wop* is the most commonly used of the three epithets in the corpus (see Table 1).

Table 1: Distribution of the hits of *wop*, *eyetie* and *goombah* in different subcorpora

RACIST EPITHETS				
		WOP	EYETIE	GOOMBAH
Total number of hits		163	10	42
METADATA				
Additional language(s) in the movie	yes	64	2	15
	no	99	8	27
Film release date	before 1970	4	1	0
	1970-1979	10	1	3
	1980-1989	15	0	1
	1990-1999	34	0	15
	2000-2009	32	3	9
	2010-2017	68	5	14
Countries the movie was shot in	Canada	9	2	0
	France	1	0	0
	Switzerland	2	0	0



	UK	13	2	0
	UKFranceCzechRepublic	1	0	0
	USA	89	1	28
	USAAustralia	2	0	2
	USACanada	1	0	1
	USACanadaGermany	1	0	0
	USAFrance	2	0	0
	USAFranceBelgium	1	0	0
	USAGermany	3	0	0
	USAItaly	2	0	0
	USAJapan	4	0	0
	USANorway	0	0	1
	USAUK	9	1	0
	unknown	23	4	10
Genre of the movie	Action	18	0	9
	Adventure	3	0	1
	Animation	1	0	1
	Biography	23	1	5
	Comedy	29	2	17
	Crime	88	4	24
	Documentary	1	0	0
	Drama	129	8	26
	Family	3	0	1
	Film-Noir	1	0	0
	History	9	1	0
	Horror	9	0	1
	Musical	4	0	0
	Mystery	10	1	1
	Romance	29	0	3
	Sci-Fi	10	0	2
	Sport	2	1	1
	Thriller	17	0	4
	War	13	2	0

Source: Author

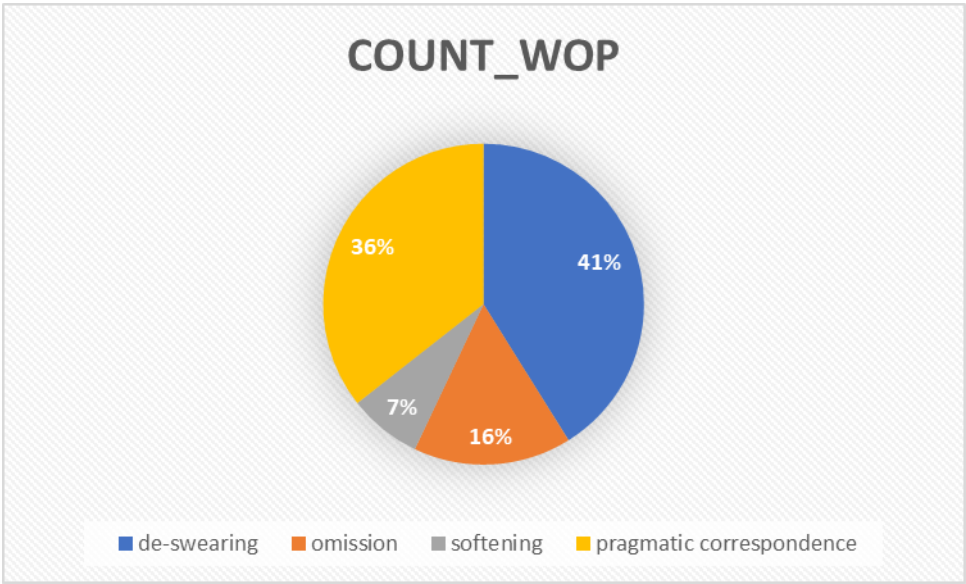
If the languages of the movies are taken into account, there are one or more additional languages in a minority of the audiovisual products where the racist epithets were found (38%, 81 out of 215 subtitle hits). As regards the countries the movies were shot in, the vast majority of slurs were found in movies set in the USA (55%, 118 out of 215 subtitle hits). The genres of the movies where the racial slurs were found are mainly drama and crime.

The hypothesis that newer films could potentially contain a higher number of racist epithets was verified by creating subcorpora by date range with *Sketch Engine* and then calculating the hits per million tokens of the three racial slurs. Hits per million tokens (HPMT) is a normalized measure of frequency that allows comparisons between corpora of different sizes. In this case, it indicates the number of times that a racial slur (*wop*, *eyetie* and *goombah*) would occur in one million words of the corpus. The HPMT found were the following: 0.36 (before 1970), 0.61 (1970-1979), 0.35 (1980-1989), 0.33 (1990-1999), 0.09 (2000-2009) and 0.13 (210-2017).

In order to analyze the renditions of the epithets, the translations were divided into the four categories adopted from Díaz-Pérez (2020). The count was carried out both separately for every racial slur taken into consideration and also for all the data collected. The count is expressed in percentages because the number of hits per epithet differs and percentages allow for a comparison

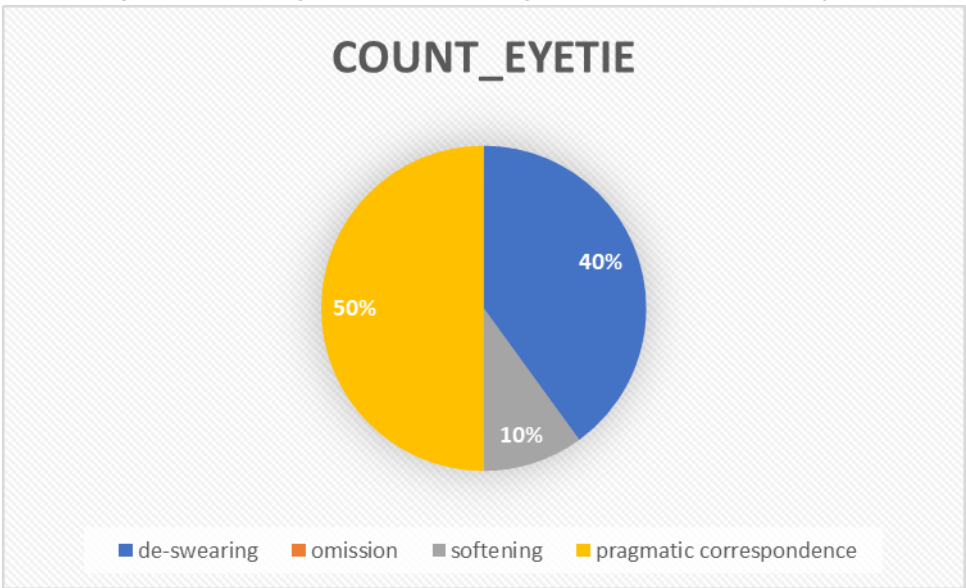
between datasets of different sizes. The results are represented graphically with pie charts (Figures 1-4).

Figure 1: Percentage of translation categories of the renditions of *wop*



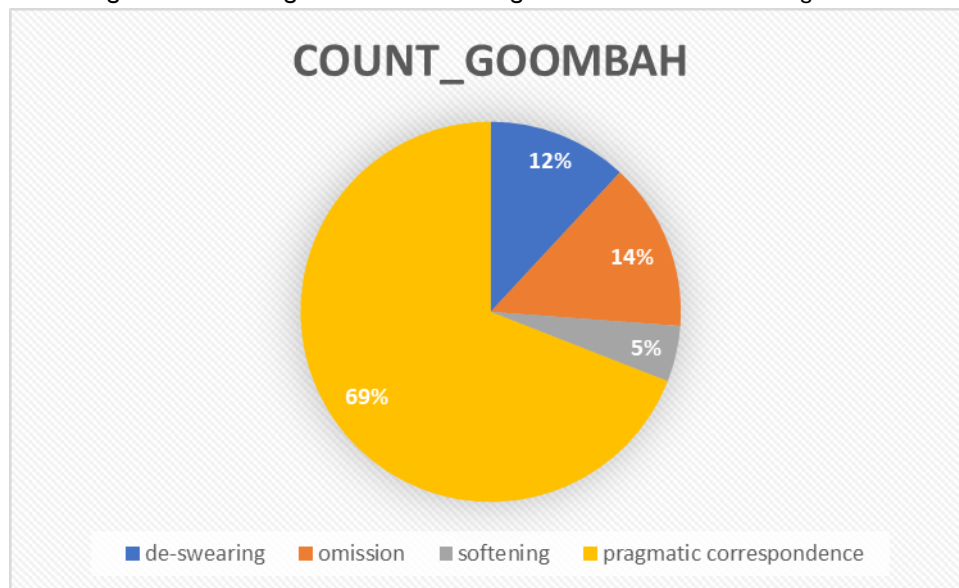
Source: Author

Figure 2: Percentage of translation categories of the renditions of *eyetie*



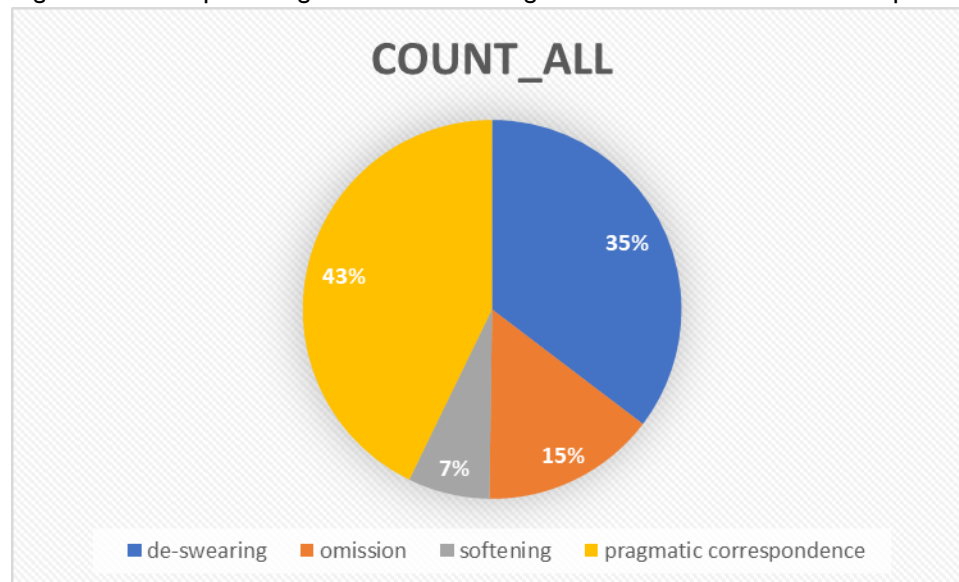
Source: Author

Figure 3: Percentage of translation categories of the renditions of *goombah*



Source: Author

Figure 4: Overall percentage of translation categories of the renditions of racist epithets

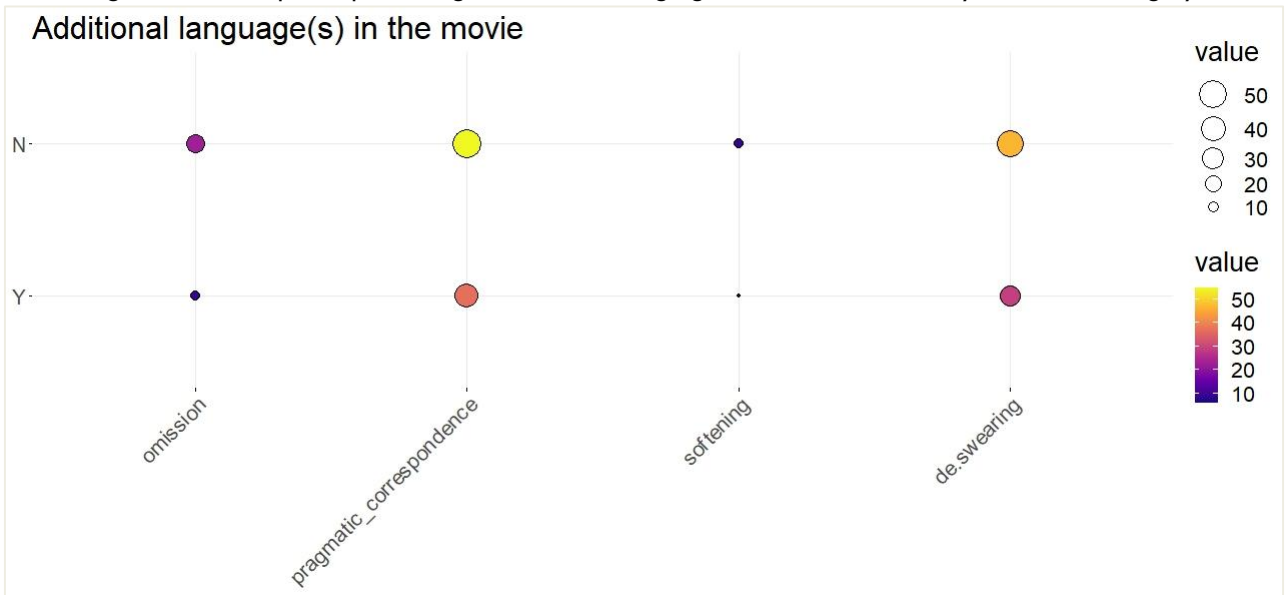


Source: Author

Overall (Figure 4), the most applied translation strategy was *pragmatic correspondence* (43%), followed by *de-swearing* (35%), whereas *softening* was the least used strategy (7%). Also *omission* was not commonly used (17%). A closer look at the three different epithets (Figures 1-3) highlights that *wop* is the epithet for which *pragmatic correspondence* was less used (36%). The frequency of this category is in fact behind *de-swearing* in the data for *wop* (Figure 1). *Goombah* is, instead, a term for which *pragmatic correspondence* was used in the majority of translations (69%, Figure 3). *Eyetie* was never omitted, differently for *goombah*, for which *omission* was the second most used option, even if far behind *pragmatic correspondence* (14%, Figure 3). The use of *de-swearing* was similar in *wop* and *eyetie* (41% and 40% respectively), while for *goombah* it covered only 12% of the translations.

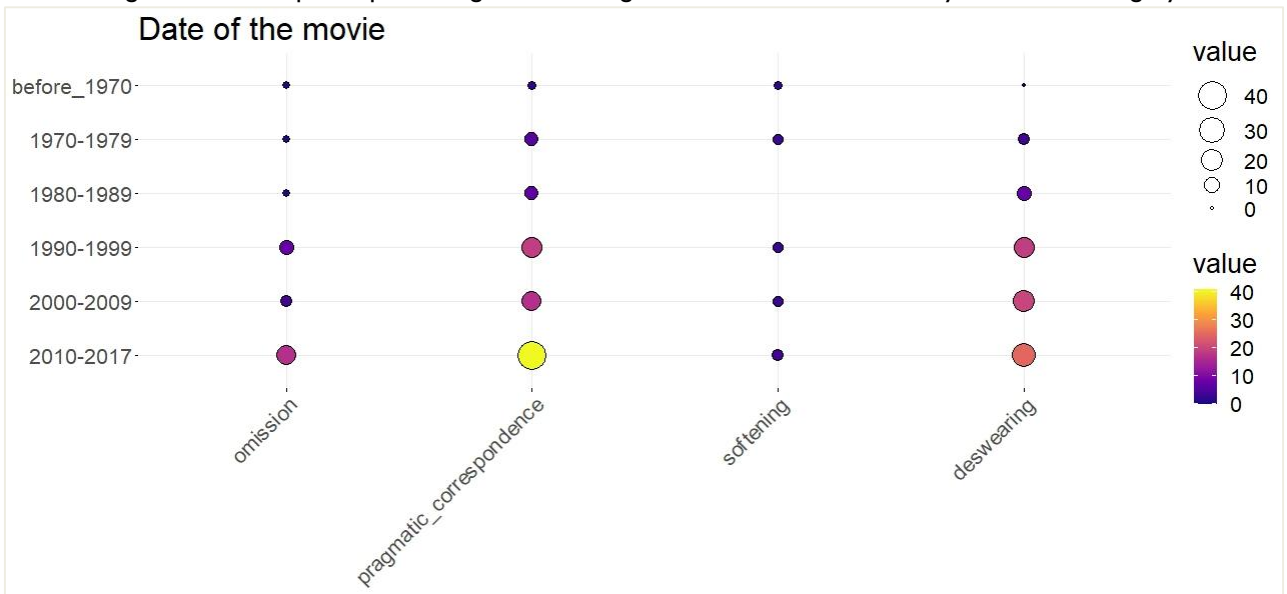
The balloon plots represented in Figures 5-8 show the distribution of translation choices divided by the presence of additional languages (Figure 5), the film release date (Figure 6), the countries the movie was shot in (Figure 7) and the genre of the movie (Figure 8).

Figure 5: Balloon plot representing the additional language in the movie divided by translation category



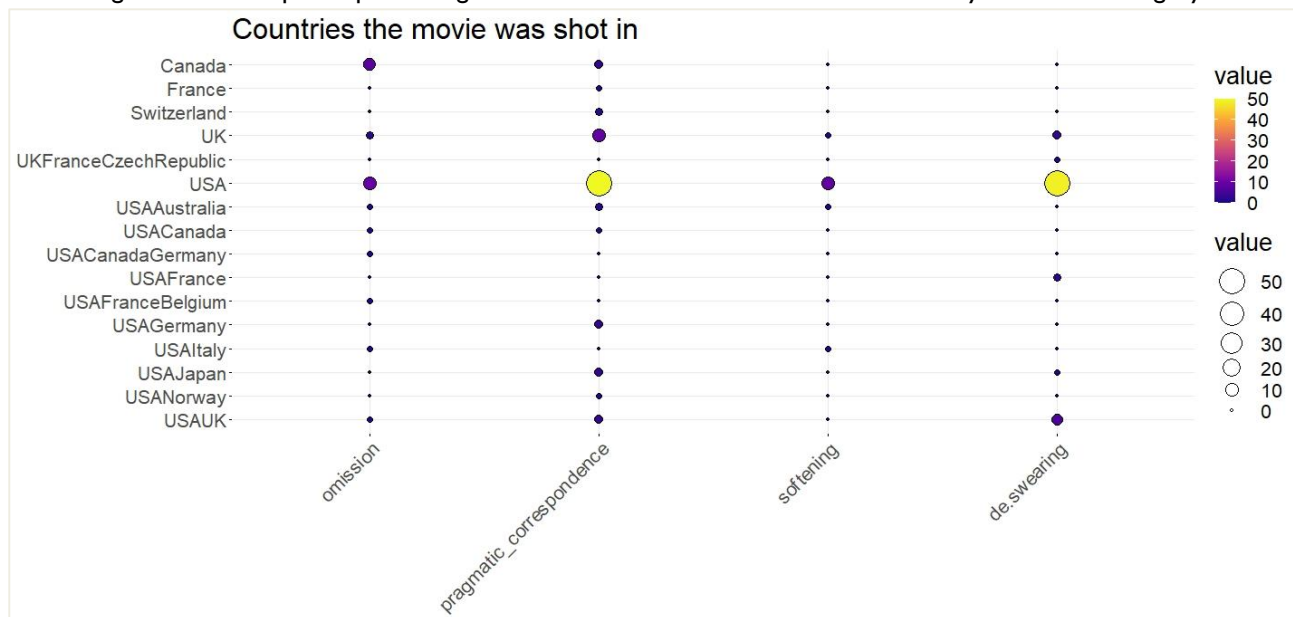
Source: Author

Figure 6: Balloon plot representing the date range film release dates divided by translation category



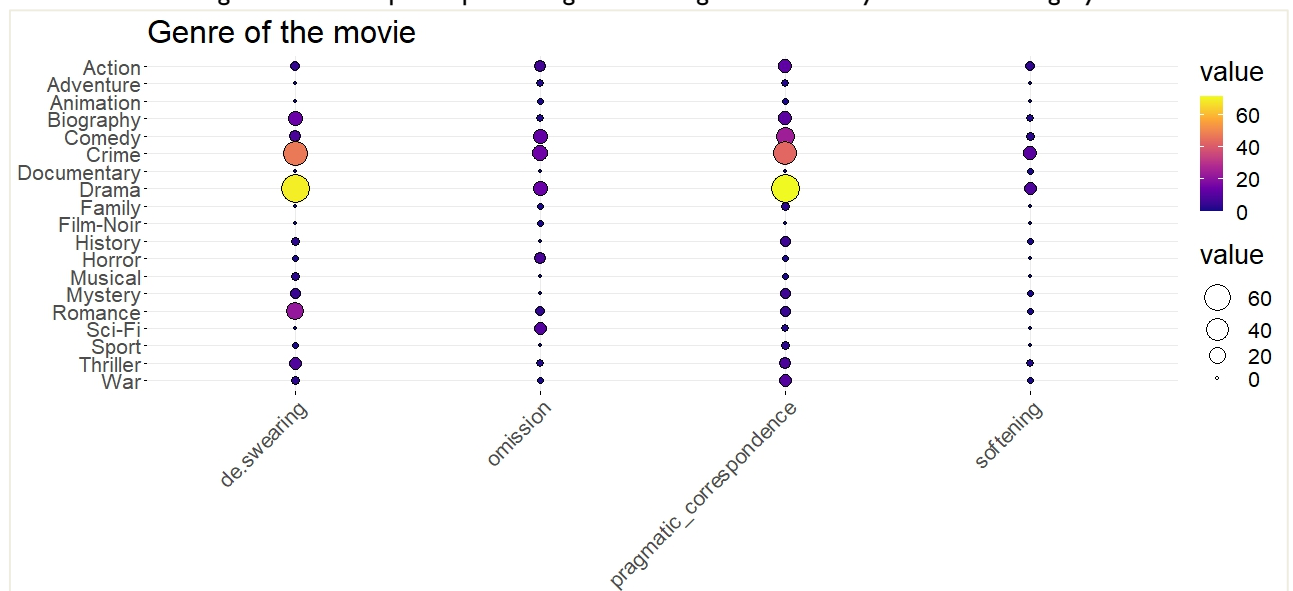
Source: Author

Figure 7: Balloon plot representing the countries the movie was shot in divided by translation category



Source: Author

Figure 8: Balloon plot representing the movie genre divided by translation category



Source: Author

In order to study the influence of the metadata variables on the strategies applied for the translation of slurs, chi-square tests with Monte Carlo simulation were conducted between the translation categories and every independent variable. The same procedure was applied also to check whether the translation strategies adopted differed significantly depending on the epithet to be translated.

There was no statistically significant association between the translation category and the presence of an additional language ($\chi^2= 1.54$, $p= 0.65$, $V= 0.00$). There was a statistically significant and moderately strong association between the translation category and the film release date ($\chi^2= 199.28$, $p= 0.00$, $V= 0.35$), between the translation category and the countries where the movie was

shot ($\chi^2= 82.92$, $p= 0.00$, $V= 0.23$) and between the translation category and the genre of the movie ($\chi^2= 163.02$, $p= 0.00$, $V= 0.26$). Eventually, a statistically significant association with a small effect was found between the translation strategies and the type of racial slur translated ($\chi^2= 19.07$, $p= 0.00$, $V= 0.17$).

7. Discussion

In the data for the three epithets, the categories *pragmatic correspondence* and *softening* represent 50% of the translations and *omission* and *de-swearing* the other 50%. Consequently, no trend to censor the racist epithets can be seen in this dataset. The overall result is influenced by *eyetie* and *goombah*, which are both rendered with the same or a similar tone in most cases. The data about *eyetie* have to be taken with caution because the number of hits in the corpus is very small and much smaller than that of *wop* and *goombah*. A focus on *wop* shows a slightly different translation approach. *Pragmatic correspondence* and *softening* are often used (43% of cases), but they do not cover the majority of translations. In 57% of subtitles where *wop* is used, the translator either tones down or omits the racist epithet. The findings about *wop* are in line with a sanitizing tendency towards offensive language in subtitle translation (Díaz-Pérez, 2020). On the one hand the translation categories significantly differ depending on the epithet and, on the other hand, a higher number of *pragmatic correspondences* were found both in the translations of *eyetie* and of *goombah*. A possible explanation of these results could be the lower offensive meaning that *eyetie* and *goombah* convey, because of which the translator does not consider sanitizing necessary.

The distribution of translation categories is not influenced by the time range when the film was shot, but a moderately strong association between the translation category and the film release date was found, with a higher number of hits classified as *pragmatic correspondence* and *de-swearing* compared to *softening* and *omission*. Moreover, the number of hits found increases as the date range becomes more recent. This could depend on a reduced influence of censorship both in the original and in the translated version of more recent movies, but it is also a consequence of a much higher number of tokens for more recent audiovisual products than for older ones. When HPMT, a normalized measure of frequency, is considered, it results that, differently from the initial hypothesis, more recent audiovisual products use fewer racist epithets than older ones. The frequency of use of racist epithets is at its highest in the seventies (0.61 HPMT), it remains stable in the eighties (0.35 HPMT) and in the nineties (0.33 HPMT) and it goes down in the periods 2000-2009 (0.09 HPMT) and 2010-2017 (0.13 HPMT). For the years before 1970 there are just 5 hits and the HPMT of racial slurs is 0.36. A possible explanation of this finding is that, even if racist epithets would not be censored, they are perceived as disturbing by the contemporary public and, therefore, avoided. Results about use in context of these racial epithets show that the large majority of them is found in movies set in the US. This finding could be explained considering that racist epithets are used against people of a minority group in a specific society (Hughes, 2006) and the US is a country where many Italians emigrated. In addition, the US is one of the major movie producers in the world. According to the data available until 2017, the US generated the third-highest quantity of movies among all the national film industries, trailing only behind India and China (UNESCO, 2023).



Therefore, the statistically significant and moderately strong association between translation category and countries where the movie was shot can be interpreted as a consequence of the overwhelming majority of films set in the USA compared to any other country. Epithets found in movies shot in multiple countries were a minority (31 out of 215) divided into 11 options, 10 of which include the US among the countries. The hypothesis that shooting in multiple countries could influence translation choices was not confirmed. It was also hypothesized that the presence of more than one language in a movie could have an impact on the translation of racist epithets, but the number of languages used does not have an impact on subtitles.

When looking at genre, the majority of hits comes from film of the genres *Drama* and *Crime*. This is in line with the kind of words looked up, because racist epithets are related to the negative stereotype of Italians as illegal immigrants associated with organized crime (Chapman, 1986; Hughes, 2006; Train, 1912).

8. Conclusion

The corpus *OpenSubtitles 2018 parallel - English* accessible through *Sketch Engine* that was used in the present study is a valuable research tool but has also limitations. The positive aspects are that it is a large collection of real subtitles (about 1.2 billion tokens) and it is very easy to query through *Sketch Engine*. In addition, not all search results had all the metadata, some of them were missing. Knowing the title of the film, it was possible to retrieve the other metadata, but when the title was missing it was impossible. In particular, for 37 hits of the original dataset there were the metadata about language, genre and date, but not about the country and, since also the film title was missing, it was not possible to retrieve this information. These data had to be discarded in the analyses about the association between translation categories and country the film was shot in. Furthermore, the information provided in the corpus about place and time focus on the movie production instead of on the story setting. Country and date refer to the countries and the year of movie shooting, not to the place and time of the events filmed. Metadata focusing on the story told in the movie, instead of on its production process, would be interesting to analyze.

The present study has performed a quantitative analysis and discussed general trends by means of descriptive and statistical analysis of the data collected. Overall, the data about *wop*, *eyetie* and *goombah* collected in *OpenSubtitles* show that these racist epithets are used more frequently in audiovisual products of genre *Drama* and *Crime* shot in the seventies in the USA. The presence of other languages in the movie is not associated with the translation strategy used to translate racial slurs into Italian. Overall, the distribution of translation categories remains constant in the different subset of data divided by the independent variables (additional language, date, country and genre). *Pragmatic correspondence* is the most used strategy, followed by *de-swearing*, *omission* and *softening*. Among the three racist epithets, *wop* is the more offensive one and this emerges also in a sanitizing tendency in its translation, which however is not present for *eyetie* and *goombah*.

The data analyzed did not allow for a more qualitative evaluation of the use of racial slurs in context. A potential research development could be a qualitative analysis informed by the quantitative results. For example, a focus on the use and translation of racist epithets in movies of



genre *Drama* and *Crime* shot in the seventies in the USA could give further insights on the meaning and pragmatic nuances that these racist epithets have in movie subtitles.

References

- Allen, I. L. (1983). *The Language of Ethnic Conflict*. Columbia University Press.
- Ávila-Cabrera, J. (2016). The Treatment of Offensive and Taboo Terms in the Subtitling of Reservoir Dogs into Spanish. *Trans*, (20), 25–40. <https://doi.org/10.24310/TRANS.2016.v0i20.3145>
- Beseghi, M. (2016). WTF! Taboo Language in TV Series: An Analysis of Professional and Amateur Translation. *Altre Modernità*, (n. spe), 215–231. <https://doi.org/10.13130/2035-7680/6859>
- Bobbitt, Z. (2023). *Statology*. Statology. <https://www.statology.org/>
- Chamizo Domínguez, P. J. (2018). Problems Translating Tabooed Words from Source to Target Language. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp. 199–217). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198808190.013.11>
- Chapman, R. L. (1986). *New Dictionary of American Slang*. Harper & Row.
- Delgado, R., & Stefancic, J. (2004). *Understanding Words That Wound*. Routledge.
- Díaz-Pérez, F. J. (2020). Translating Swear Words from English into Galician in Film Subtitles: A Corpus-based Study. *Babel*, 66(3), 393–419. <https://doi.org/10.1075/babel.00162.dia>
- Filmer, D. (2012). ‘The “Gook” Goes “Gay”’—Cultural Interference in Translating Offensive Language. *InTRAlinea*, 14. <https://www.intraline.org/archive/article/1829>
- Foley, M. (2023). 6.7 Case Study: Chi-Square, Fisher. *Statistical Inference: Data Analyst Handbook*. <https://bookdown.org/mpfoley1973/statistics/case-study-chi-square-fisher.html>
- Gottlieb, H. (1992). Subtitling – A New University Discipline. In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting. Training, Talent and Experience* (pp. 161–170). John Benjamins.
- Greenall, A. K. (2011). The Non-translation of Swearing in Subtitling: Loss of Social Implicature? In A. Serban, A. Matamala & J.-M. Lavour (Eds.), *Audiovisual Translation in Close-up: Practical and Theoretical Approaches* (pp. 45–60). Peter Lang.
- Hogg, M. A., & Abrams, D. (1998). *Social Identifications – A Social Psychology of Intergroup Relations and Group Processes*. Routledge.
- Hope, A. C. A. (1968). A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3), 582–598. <https://doi.org/10.1111/j.2517-6161.1968.tb00759.x>
- Howitt, D., & Cramer, D. (2020). *Introduzione alla statistica per psicologia*. M. Benassi, R. Bolzani & G. Rossi (Eds.). Pearson.
- Hughes, G. (2006). *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-speaking World*. Routledge.
- Jain, N., & Warnes, G. R. (2006). Balloon Plot – Graphical Tool for Displaying Tabular Data. *R News - The Newsletter of the R Project*, 6(2), 35–38.
- Kelley, K., & Preacher, K. J. (2012). On Effect Size. *Psychological Methods*, 17(2), 137–152. <https://doi.org/10.1037/a0028086>



- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography*, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 923–929). European Language Resources Association.
- Memmi, A. (1999). *Racism*. University of Minnesota Press.
- Merriam-Webster. (n.d.). Goombah. In *Merriam-Webster Dictionary*. <https://www.merriam-webster.com/dictionary/goombah>
- Reisigl, M., & Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. Routledge.
- RStudio, R. C. T. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (2022.12.0) [Computer software]. <https://rstudio.com/>
- Soler Pardo, B. (2015). *On the Translation of Swearing into Spanish: Quentin Tarantino from Reservoir Dogs to Inglourious Basterds*. Cambridge Scholars Publishing.
- Sotelo, P. (2015). Design and Applications of the Veiga Corpus: Using a Multimedia Corpus of Subtitles in Translation Training. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (pp. 245–266). John Benjamins. <https://doi.org/10.1075/scl.69.12sot>
- The University of Utah. (2023). Chi-Square – Sociology 3112. Department of Sociology. *The University of Utah*. <https://soc.utah.edu/sociology3112/chi-square.php>
- Tiedemann, J. (2009). News from OPUS — A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov, G. Angelova & R. Mitkov (Eds.), *Recent Advances in Natural Language Processing V: Selected Papers from RANLP 2007* (pp. 237–248). John Benjamins. <https://doi.org/10.1075/cilt.309.19tie>
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In N. Calzolari et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 2214–2218). European Language Resources Association.
- Torres Cuenca, A. (2016). *On the Translation of Taboo Words in an English-Spanish Corpus of Film Subtitles*. [Dissertation]. Universidad de Jaén.
- Train, A. C. (1912). *Courts and Criminals*. Library of Alexandria.
- Treccani. (n.d.). Compare. In *Vocabolario Treccani*. <https://www.treccani.it/vocabolario/compare>
- UNESCO. (2023). UIS Statistics. UNESCO Institute for Statistics. <http://data.uis.unesco.org/?ReportId=5538>

Notes

Authorship contribution

Conceptualization: S. Ghiselli

Data collection: S. Ghiselli

Data analysis: S. Ghiselli



Cadernos de Tradução, 44(Special Issue 2), 2024, e99464
Graduate Program in Translation Studies
Federal University of Santa Catarina, Brazil. ISSN 2175-7968
DOI <https://doi.org/10.5007/2175-7968.2024.e99464>

Results and discussion: S. Ghiselli

Review and editing: S. Ghiselli

Research dataset

The data were extracted from the corpus *OpenSubtitles 2018 parallel - English*, available inside the platform Sketch Engine (I used the access provided by the University of Bologna). The racist epithets extracted and analyzed were not used in other projects but extracted, exclusively, to write this article.

Funding

Not applicable.

Image copyright

Not applicable.

Approval by ethics committee

Not applicable.

Conflicts of interest

Not applicable.

Data availability statement

The data from this research, which are not included in this work, may be made available by the author upon request.

License

The authors grant *Cadernos de Tradução* exclusive rights for first publication, while simultaneously licensing the work under the Creative Commons Attribution ([CC BY](#)) 4.0 International License. This license enables third parties to remix, adapt, and create from the published work, while giving proper credit to the authors and acknowledging the initial publication in this journal. Authors are permitted to enter into additional agreements separately for the non-exclusive distribution of the published version of the work in this journal. This may include publishing it in an institutional repository, on a personal website, on academic social networks, publishing a translation, or republishing the work as a book chapter, all with due recognition of authorship and first publication in this journal.

Publisher

Cadernos de Tradução is a publication of the Graduate Program in Translation Studies at the Federal University of Santa Catarina. The journal *Cadernos de Tradução* is hosted by the [Portal de Periódicos UFSC](#). The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

Guest editor

Gian Luigi De Rosa

Section editors

Andréia Guerini – Willian Moura

Technical editing

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Article history

Received: 06-04-2024

Approved: 10-05-2024

Revised: 06-06-2024

Published: 06-2024



Cadernos de Tradução, 44(Special Issue 2), 2024, e99464
Graduate Program in Translation Studies
Federal University of Santa Catarina, Brazil. ISSN 2175-7968
DOI <https://doi.org/10.5007/2175-7968.2024.e99464>



Available in:

<https://www.redalyc.org/articulo.oa?id=731981875010>

How to cite

Complete issue

More information about this article

Journal's webpage in redalyc.org

Scientific Information System Redalyc
Diamond Open Access scientific journal network
Non-commercial open infrastructure owned by academia

Serena Ghiselli

**A quantitative analysis of racist epithets referring to
Italians and their translations in movie subtitles: The
case of wop, eyetie and goombah**

Cadernos de Tradução

vol. 44, no. 2, Esp. e99464, 2024

Universidade Federal de Santa Catarina,

ISSN: 1414-526X

ISSN-E: 2175-7968

DOI: <https://doi.org/10.5007/2175-7968.2024.e99464>