

Chapter 3

Racism goes to the movies: A corpus-driven study of cross-linguistic racist discourse annotation and translation analysis

Effie Mouka, Ioannis E. Saridakis and Angeliki Fotopoulou

This paper traces register shifts (Halliday & Hasan 1976: 22; Hatim & Mason 1997) between source-texts (English) and target-texts (Greek and Spanish) in instances of racist discourse in films. It presents preliminary, as yet non-exhaustive, findings and aims to ultimately formulate explanatory hypotheses concerning the emerging norms. Our methodological approach is placed in the framework of Descriptive Translation Studies (Toury 2012; Chesterman 2008) and in the school of Critical Discourse Analysis (Fairclough 1985; 1992), relying on Appraisal Theory (Martin & White 2005) to provide and analyse a taxonomy of the racism-related utterances examined.

1 Introduction

Technological advances in Corpus Linguistics and tools for processing and compiling linguistic corpora open new ways on how we exploit textual and research material. In a descriptive approach, textual and pragmatic annotation can largely facilitate the systematic lexico-grammatical analysis of linguistic resources (see McEnery & Hardie 2012: 29–31; Zanettin 2012: 76–79). This holds true also for translation corpora, with a particular focus on the descriptive examination of translation strategies and norms (Zanettin 2012: 78–96).



Effie Mouka, Ioannis E. Saridakis & Angeliki Fotopoulou. 2015. Racism goes to the movies: a corpus-driven study of cross-linguistic racist discourse annotation and translation analysis. In Claudio Fantinuoli & Federico Zanettin (eds.), *New directions in corpus-based translation studies*, 35–70. Berlin: Language Science Press

This paper partly presents the first author's¹ ongoing PhD research, which aims to examine, from a descriptive viewpoint and by using corpus annotation, the translational norms of the socio-culturally marked discourse of racism, and the shifts observed during the discourse transfer from a source language (EN) into two target languages (EL, ES).² This paper focuses on the applied methodology, on the findings collected so far, and discusses problems and impediments observed during corpus analysis.

Racism, as manifested in discourse, is a constantly open issue that merits research (van Dijk 1993; Reisigl & Wodak 2001) and is clearly on the agenda of (critical) discourse analysis in light of the European social, political and economic backdrop. Realistic films on racism represent discourses emanating from racist stances, while cinema, as a medium widely accessible to the public communicates ideas apart from reflecting society. On the other hand, subtitles are considered to be among the most read translations and text types in countries with a subtitling tradition (Gottlieb 1997: 153 in Pedersen 2011: 125). To this end, the analysis of subtitles in racism-related films, rather than in films with sporadic racist utterances, seems to be better suited to research on the translation of racist-oriented discourse.

§2 of this article outlines the aims and scope of our research, and introduces the basic concepts and theoretical tenets used in this study. First, we introduce the principal discourse-related definitions of racism, together with a discussion of how racist discourse is handled by Critical Discourse Analysis. Subsequently, we provide a brief overview of Appraisal Theory, first developed by Martin & White (2005). This theory has been used extensively in sentiment analysis. Finally, we consider the phenomenon of register shifts in subtitles.

§3 presents our corpus-driven methodology and the corpus tools used in our research. §4 and §5 present and exemplify the implementation of our methodology and outline the principal findings with regard to context-bound register shifts in translation.

¹ The second author, I.E. Saridakis is the PhD research director. Dr. A. Fotopoulou also participates in the project's consultative committee. The authors express their gratitude to V. Giouli, scientific associate at the Institute for Language and Speech Processing (ILSP) for her support in initially developing and implementing the sentiment annotation scheme described in this paper, and in adopting the corpus metadata handling model used in our method.

² Batsalia & Sella-Mazi (2010: 120–121) define “shifts” as subsuming all changes that may appear during the translation process, on a semantic, lexical, morphological, syntactic, pragmatic, and/or stylistic level. The “translation shift” hypothesis is a useful and powerful descriptive device, to approach hermeneutically the phenomenon of differentiation of the TT from its ST, without stigmatising it.

2 Research aims and scope

The focus of our work is to examine racist discourse from a translation perspective, identifying its structure, its textual deployment, and its elements and traits on the basis of lexicogrammatical evidence and using a classificatory device. In other words, our aim is to examine how racist attitudes can be classified in spoken film discourse, linking this classification to the context of the utterances from which the text chunks have been drawn. This classification and analysis is based on a model adapted from Appraisal Theory (Martin & White 2005), using postulates derived from Critical Discourse Analysis (Reisigl & Wodak 2001; van Dijk 2000a; 2000b; 2002). Finally, by linking the examined ST utterances to their translations in two T_Ls, register shifts can be analysed on the basis of previous research (Hatim & Mason 1997; Mason 2001; Pettit 2005; Mubenga 2009; Munday 2012). This study is based on corpus resources and methodologies. We first constructed an ad hoc corpus and annotated it with a purpose-built annotation scheme, then set out to identify register shifts in the translation of racist utterances. This approach is exemplified by the preliminary findings reported in this article.

2.1 Background. Racism and racist discourse

The phenomenon of racism is fuzzy and evasive, and the term is often used rather vaguely, even to describe discriminatory phenomena other than those related to the concept of “race”. *Racism* subsumes everyday practices and behaviours, both verbal and non-verbal, stereotyping, discriminatory practices, institutional systemic policies, or even acts of racial segregation and genocides (Giddens 2009: 637–653).

How racism is defined depends, in the final analysis, on the scope of individual research: for example, literature lists distinctive definitions such as “institutional” or “systemic” racism, to designate racism that is present in societal structures, such as the educational system or the police; “neo-racism” or “cultural racism” that draws from cultural differences in an attempt to provide explanations for inequalities and the actual position of ethnic minorities, immigrants and refugees in society, as opposed to the “old-style” and merely biologically explained racism that is based on physical characteristics to sustain the inferiority of certain group members; “everyday racism” as a common societal behaviour; or racism as part of the wider phenomenon of “heterophobia”, the fear of the Other, that gives birth to various forms of discrimination (Essed 1991; Reisigl & Wodak 2001; Memmi 2000: 118). Racism has “the cognitive function of organizing the social represen-

tations (attitudes, knowledge) of the group, and thus of indirectly monitor[ing] the group-related social practices, and hence also the text and talk of members” (van Dijk 1995: 248).

We share Van Dijk’s position that ideologies are systems with both a cognitive and a social dimension, in other words “belief systems” that involve ideas, judgments, values and attitudes shared by members of social groups and targeting other social groups.

2.2 Racism in Critical Discourse Analysis and Corpus Linguistics

Racist discourse has been investigated mainly within the framework of Critical Discourse Analysis (CDA). It has been shown that racist discourse *about* and *addressed at* minorities and immigrants tends to use the following means: lexicon and especially referential and predicative strategies; syntax, i.e. use of passive instead of active voice; rhetorical devices such as metaphors, metonymies and connotations (*synecdoches*); argument schemata; pragmatic features; recurrent topics concerning differences mainly in terms of habits, culture, language or religion of others, or even representations of others as a threat for “our” jobs, safety and culture; standard arguments and fallacies; and local moves or intensifying and mitigation strategies (Reisigl & Wodak 2001; van Dijk 2000a; 2000b; 2002).

In a study developing along lines that are ideationally and hermeneutically analogous to the study reported in this paper, Baker et al. (2008) have effectively used a Corpus Linguistics approach to analyse a large corpus of news articles about refugees, asylum seekers, immigrants and migrants and have shown that such an empirical approach can fruitfully combine CL and CDA.³

2.3 Appraisal Theory and Sentiment Analysis

Appraisal Theory is a model that has evolved within the theoretical framework of Systemic Functional Linguistics. It “is concerned with the interpersonal in language, with the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate [...], with how writers/speakers approve and disapprove, enthuse and abhor, applaud and criticise, and with how they position their readers/listeners

³ Research in the field of CADS, or Corpus-Assisted Discourse Studies focuses on the use of corpora and corpus analysis techniques, so as to unveil meaning and style in discourse and to examine particular discourses. For a comprehensive bibliography on CADS compiled by C. Gabrielatos, see: <http://goo.gl/WHB2mh>.

to do likewise” (Martin & White 2005: 1). In addition, appraisal “co-articulates interpersonal meaning”⁴ (Martin & White 2005: 33), reflecting the speakers’ social roles and interpersonal positioning, and their inter-subjective negotiation. It is an approach that can be used to explore, describe and explain how language is used to express stances.⁵

The Appraisal framework has been widely used in Sentiment Analysis to identify subjective information, emotions and opinions, as manifested in discourse (see e.g. Whitelaw, Garg & Argamon 2005; Taboada & Grieve 2004; Asher, Benamara & Mathieu 2009) and to classify the attitude of speakers/writers. More recently, this type of analysis has been applied in translation research (see the work reported in Munday 2012: 42–79), with the aim to “describe the different components of a reader’s attitude, the strength of that attitude (graduation) and the ways that the speaker aligns him/herself with the sources of attitude and with the receiver (engagement)” (Munday 2012: 2).

The three sub-components of Appraisal (theory) are *attitude*, *graduation* and *engagement*. *Attitude* is concerned with affect, judgement and appreciation and has a polarity, i.e., a positive or a negative dimension. *Engagement* deals with the positioning of the speaker towards the evaluation and concerns the rhetorical devices that are used to vary the engagement of speakers with their utterances (*I believe...*, *it is rumoured that...*, *X said...*). *Graduation* concerns grading phenomena and adjusting the degree of evaluations (e.g., in the grading between competent player, good player, brilliant player or contentedly, happily, joyously, ecstatically). Moreover, graduation is applicable also to indicators of engagement. A bare assertion does not have the same intensity as an utterance that is introduced with a modal value, such as *possibly* or *certainly* or presented as a hypothesis (compare the following sentences as to the grade of engagement they

⁴ According to Halliday (1978: 112), “the interpersonal component represents the speaker’s meaning potential as an intruder. It is the participatory function of language, language as doing something”. Through the interpersonal component, the speaker places himself in the context of the situation, to express also “his own attitudes and judgements” and thus to seek to “influence the attitudes and behaviours of others” (ibid.). In other words, the interpersonal meaning can be defined as the one deriving from the roles and relationships, e.g. status, intimacy, contact, sharedness between interactants (Egins & Slade 1997: 49). Finally, for Kress (2010: 94), meaning-making is “both social and external and social and internal [...] Meaning is constantly created in a transformative process of interactions with and in response to the prompts of social others and of the culturally shaped environment; and there is constant ‘internal’ action, an (inner) response in constant engagement with the world”.

⁵ In this context, *stance* is meant to represent “performances through which speakers may align or disalign themselves with and/or ironize stereotypical associations with particular linguistic forms” (Jaffe 2009: 4).

represent: *They're all a bunch of fucking freeloaders. Some of them are all right I guess*) (see Martin & White 2005: 136 for detailed examples of how graduation applies to attitudinal meanings and engagement values).

2.4 Subtitling: Analysis of register shifts

The ultimate aim of this study is to analyse register shifts in subtitled films. Munday suggests that “*major shifts in key attitudinal markers are not likely to occur except perhaps in certain genres [...] where the TL conventions are strongest*” (Munday 2012: 159–160; our emphasis) and calls for the inclusion of Appraisal Theory in the study of how the interpersonal meaning is materialised in translation (ibid.). Munday regards evaluative expressions in text as critical points that pose problems for translators/interpreters and which are especially susceptible to translators’ interventions: he investigates the way a translator’s subjective stance is manifested by shifts in translations and concludes that translation choices indicate “both an ideological and axiological position” (Munday 2012: 155). The translator is a critical reader of the original, and he/she is likely either to reproduce the ideological stance of the source text by complying with its general orientation, oppose it by resisting its ideological postulates, or to both reproduce and rework the source in a more tactical approach that repositions the audience in relation to the writer (Munday 2012: 157). Such stance-taking attitudes on the part of the translator–as–critical–reader depend on the overall evaluative prosody of the text, the value of which “the translator sometimes feels obliged to explicate [...] in the interests of the target audience” (ibid.). On a more theoretical line of thought, Batsalia & Sella-Mazi (2010: 215) relate register to style, and speak of stylistic shifts in translation. These are considered to be changes of the language level(s) (i.e. lexical, semantic, pragmatic or morphosyntactic) that can be mapped to stylistic choices. Such choices derive from the translator’s own perception of the potential offered by the system of the TL and of the stylistic habits governing the text genre at hand, in a given language. More in particular, subtitling research can optimally incorporate discourse analysis techniques, so as to depict the mechanism through which racism is reflected on cinematographic discourse.

The interpersonal dimension of discourse has been explored using various models in subtitling research. Hatim & Mason (1997: 78–96) and Mason (2001) have drawn attention to translation shifts that relate to the interpersonal dynamics of film characters, pointing out that tenor is the metafunction of discourse, which is frequently sacrificed in subtitles. Their analytical approach is that of interpersonal pragmatics, with a focus on politeness features and face-threatening

acts. They compare source and target texts using a qualitative method with examples from French films translated into English. On the other hand, Pettit (2005) analyses subtitled and dubbed French versions of two English films, with the aim to examine whether the registers presented in these films are maintained or discarded in the respective subtitled and dubbed versions, providing a variety of examples of formal and informal registers, including slang and sophisticated language. Finally, Mubenga (2009) proposes a multi-modal pragmatic analysis using an SFL approach, which takes into account three components (visual, functional and cognitive) and three levels of analysis (ideational, interpersonal, textual). While sharing the basic assumptions of the above-mentioned studies, our work focuses on register shifts within a specific type of discourse, i.e. racist discourse. Also, in line with Munday (2012), it relies on Appraisal Theory to classify the translators' stance-taking tendencies. We use corpus linguistic techniques to explore translation shifts, by comparing source and target texts in two language pairs.

This study first makes extensive use of linguistic annotation to analyse translational behaviour and to investigate register shifts that take place in racist discourse; subsequently, it validates the data using large monolingual reference corpora. Based on a thorough examination of concordance lines, it provides an analysis of cross-linguistic patterns and of the collocational strength of the lexemes examined.

The aim of cross-linguistic analysis is thus to observe how and to which extent translators have addressed the categories of attitudes, relying on the annotation of register shifts. Using the GATE annotation system (Cunningham et al. 2002), register shifts were classified as *neutral transfer*, *over-toning*, *under-toning*, and *possible change of register category*. This allowed for the retrieval of bilingual segments containing examples of these shifts.

3 Corpus methodology

3.1 Corpus description

For the purpose of the present research we have developed a relatively small-scale translational film corpus consisting of the transcriptions of the dialogues of five films in English and their Greek and Spanish subtitles.⁶

⁶ The corpus forms part of the Metashare initiative launched by ELDA (see: metashare.elda.org). Details on the linguistic resource (Multilingual aligned corpus of subtitles annotated for sentiment) can be found at <http://goo.gl/o45NK5>.

The criteria used for the design of the corpus relate to the sociolinguistic situation and communicative functions (Saridakis 2010: 49–52) of the target population and can be summarised as follows:

- all films are feature films, belonging to the drama genre;
- they were all produced between 1989 and 2006 and the frame of reference of their plot is contemporary;
- their story revolves around racism and inter-racial relations;
- their approach to events and the depiction of characters have a realistic perspective;
- they contain verbally expressed racism in conversations and/or in monologues.

The five films of the corpus are:

- *Do the Right Thing* (Lee 1989);
- *American History X* (Kaye 1998);
- *Monster's Ball* (Forster 2001);
- *Crash* (Haggis 2004);
- *This is England* (Meadows 2006).

The total playtime of the corpus films corresponds to approximately 9 hours of audiovisual material, which translates into 51,000 words of transcribed English dialogue and 29,700 and 35,900 words of Greek and Spanish subtitles, respectively.

A wide variety of situations involving racist instances is present in our corpus, including both racist discourse *directed at* ethnically different Others, and racist discourse *about* ethnically different Others (van Dijk 2004: 351), which equally concern our investigation. The fact that our data contain a significant amount of conversations in inter-racial environments and everyday situations allows us to examine racist discourse *directed at* Others, and overt racist discursive practices. Such features are not always present in corpora of other genres, e.g. in political and journalistic discourse where racism has been studied more extensively

using authentic linguistic data.⁷ Cinematographic discourse can be considered as closely representing spoken impromptu racist discourse, more specifically in utterances that are *directed straight at Others*. Moreover, the communicative scenario from which our samples have been gathered is perhaps the only one permitting the construction of a parallel corpus for the cross-linguistic study of racist discourse from the perspective of translation, given that collecting samples of spoken language from other scenarios is impracticable.

Greek and Spanish were chosen as target languages because their respective cultures share some common characteristics. During the last decades, Greece and Spain have progressively seen large irregular immigration flows from non-European states, as both countries are key entry-points to Europe. Based on 2013 statistical data, non-European immigrants in Greece and Spain represented respectively 5.96 percent and 6.45 percent of the total country's population.⁸ It is noteworthy that, with the exception of Greece and Spain, only Italy experiences comparative immigration flows from non-European countries.⁹ In other words, this phenomenon has had a marked impact on the linguistic behaviour (i.e. interpersonal meaning), and therefore on racist discourse, as delineated in this study. However, although immigration is considered to have triggered racist attitudes across Europe, a social survey on racism and xenophobia has shown that the Greek population has quite stronger negative attitudes towards minorities when compared to the Spanish population.¹⁰ The phenomenon of immigration has had a marked impact on racist discourse and therefore on interpersonal meaning, as negotiated by linguistic behaviour.

A corpus-based approach facilitates the comparison of source and target segments and can help explore the extent of reflection on different socio-cultural realities and attitudes and the effect on the linguistic choices of subtitlers.

In terms of translation practices, previous work in the EL-ES language pair has *inter alia* shown that Greek subtitlers tend to omit repeated or redundant information to a more significant extent when compared to their Spanish counterparts (Sokoli 2009). In short, similar phenomena are bound to be related to generalised tendencies, and this can assist our effort in explicating the textual findings in our corpus.

⁷ E.g., in the research reported in Baker et al. (2008); van Dijk (1993), etc.

⁸ See: Eurostat 2014 data <http://goo.gl/hbouLx>.

⁹ Cmp. Vasileva (2011: 4); and <http://goo.gl/NnhjYR>.

¹⁰ See: Special Eurobarometer 2007. *Discrimination in the European Union* <http://goo.gl/t9e6Gk>.

3.1.1 Multimodality

Films are polysemiotic discourse events (Gottlieb 2005). The explication of the meaning of utterances cannot rely solely on the verbal component, but must also take into account its context: “Any consideration of film text in isolation from the context of the moving image and soundtrack is [...] doomed to failure” (Mason 2001: 23). The audiovisual text is expressed and completed through four semiotic channels (Zabalbeascoa 2008):

- audio-verbal (words uttered);
- audio-nonverbal (all other sounds);
- visual-verbal (writing);
- visual-nonverbal (all other visual signs).

Hence, there can be no valid analysis of the instances of racist discourse if the filmic message is not considered as a whole. The film provided the communicative context for each utterance and for the stance of each character which, additionally, is not static but can change as the plot unfolds. The actual use of the words and its pragmatic features can be deciphered only in this context. “Every clause [...] is a combination of ideational, interpersonal (identity and relational) and textual meanings [...] People make choices about the design and structure of their clauses which amount to choices about how to signify (and construct) social identities, social relationships, and knowledge and belief” (Fairclough 1992: 76). Even apparently innocent lexical choices can indicate a racist stance, while key lexical items are not necessarily racist: a typical example is probably the use of the word *nigger* as an indicator of *belonging* in an otherwise non-racist dialogue between black people:

“You want to get killed, nigger?”

(*Crash*, in a scene where a black male attempts an armed robbery against the driver of a luxury car and, to his surprise, the driver who resists, is also a black male).

In this case, the word itself retains its ideological profile and implications as a re-appropriated word. The connotation of key lexical words cannot be taken for granted *a priori* and out-of-context. This is because “the meaning of sentences, clauses, nouns, nominalisations and adjectives are all possible targets for the expression of ideological content, usually in the form of evaluative concepts. In

all cases, however, such a semantic representation of opinions in attitudes or models *needs to be analysed in context*" (van Dijk 1995: 260; our emphasis).

Moreover, this corpus, comprised of transcribed (i.e. originally oral) STs and written TTs, can also be characterised as a compilation of inter-semiotic facts (Jakobson 2012) that imply an inherent change of the mode of the message. Thus, certain features of the ST, such as non-standard dialects or code-switching, are not expected to be present in the TT(s). Such features are not usually (or at least, not automatically) transferred to the TL (Hatim & Mason 1997: 78). This change implies also that some prosodic features of the source dialogues are not represented in the written subtitles. This holds true also for other non-verbal features, e.g. for gestures. Such features are elements of a semiotic system (the "para-language") that interrelates with language (Halliday & Matthiessen 2014: 32–42) and have been used in our analysis in order to properly contextualise and decode the meaning of the ST. However, it would be very difficult, and well outside the scope of this study, to trace whether and up to what point the SL and TL audiences decode them in the same way:

“Modes differ in what they offer from culture to culture [...]: the different requirements of different societies and their members and the consequent different shaping. As a semiotic resource, *image* in one culture is therefore not identical to *image* in another. Even across closely related cultures and ‘languages’ (such as English, French, German) differences in the cultural use of, say, *vocal intensity* (appearing as *accent* in words and as *rhythm* in extended speech) or of *pitch variation* (appearing as intonation); differences of pace, of *vocalic quality*, and so on, lead to characteristic variation in meanings made, in signs” (Kress 2010: 81; emphasis in the original).

3.2 Corpus tools

We orthographically transcribed the English dialogues and, using ELAN (Brugman & Russell 2004), aligned all utterances to the corresponding translated Greek and Spanish subtitles. ELAN is a tool designed for complex annotations of video and audio resources. It allows the use of multiple layers of annotation, which are attributed to each speaker and are time-aligned. In other words, the transcript is aligned to the Greek and Spanish subtitles available in the DVD distributions of the films, thus making the corpus a *star corpus*, i.e. source texts in one language, and their aligned texts in multiple target languages (Johansson 2003: 140–141; Saridakis 2010: 260). As shown in Figure 1, the corpus can be expanded to include more target languages in the future, to allow for further cross-linguistic

investigations of the translational phenomena. Each utterance is accompanied by a time marker pointing to the corresponding “time-slot” in the film. ELAN allows also the inclusion of metadata to tag the speakers of individual utterances and exports text in TEI-compliant XML format using the TEI-Drop application (Schmidt 2011), and thus ensures interoperability with the applications used in subsequent stages of our research. The investigator can access the exact point of each utterance, detect the scene from which each utterance has been extracted and reconstruct the communicative context, so as to properly interpret it.

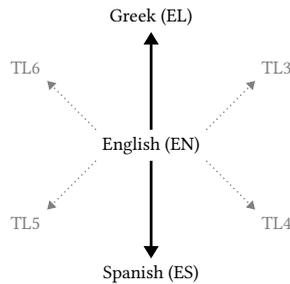


Figure 1: The “star” layout of the study corpus

Concerning the description of the corpus contents, the following metadata has been used (based on the TEI guidelines):

- for files (in the file header section): the audio and video file name/path, the film title, the publication statement (information concerning the distribution of the text);
- for utterances: information on the speaker (e.g., SPK1 [Danny]), the timestamp (*anchorsynch*), and the dependent/aligned tiers with the Greek and Spanish translations respectively (*spanGrp* type).

An example of an XML-tagged utterance and of its translations into EL and ES, as outputted from ELAN following processing in TEI-Drop, is in Figure 2.

ELAN is used to visualise each utterance in its context, together with the speaker and the aligned subtitled utterances.

To manually annotate the corpus, we have used the GATE platform (Cunningham et al. 2002).¹¹ The annotation was performed on the level of text chunks,

¹¹ The annotation methodology and a preliminary annotation scheme (regarding emotions) were first presented in Mouka et al. (2012). The annotation scheme used in this paper was later finalised and has been reported in Mouka (2014).

```

<div>
<u who="\#_LSPK1">
<anchor synch="\#_LT3497"/>
"We_are_not_enemies,_but_friends.
<anchor_synch="\# T3498"/>
</u>
<spanGrp_type="subtitles el">
<span_to="\#T 3498"_from="\# T3497">"Δεν είμαστε εχθροί , αλλά φίλοι .</
span> </spanGrp>
<spanGrp type="subtitles es">
<span to="\#T_3498" from="\#_LT3497">"No_somos_enemigos,_sino_amigos.</
span></spanGrp>
</div>

```

Figure 2: Example of an XML-tagged utterance

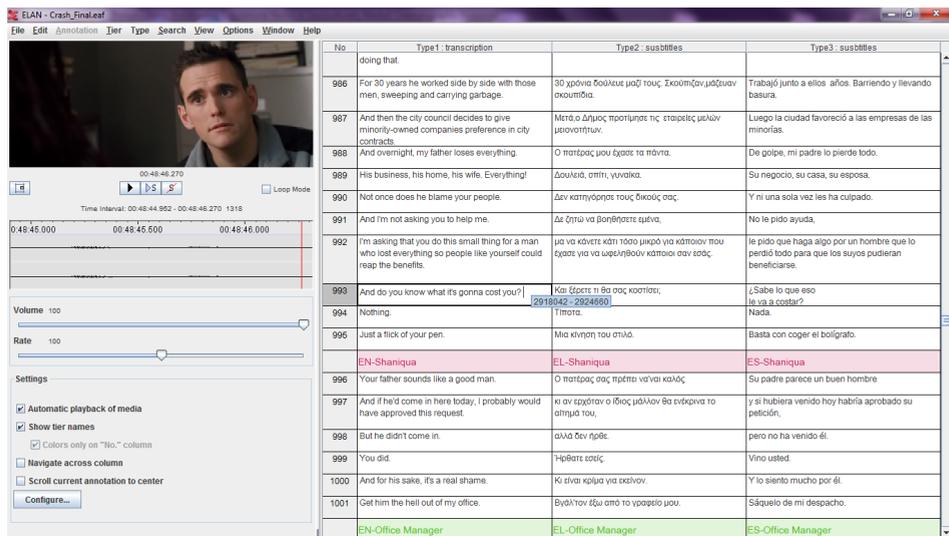


Figure 3: ELAN interface, in transcription mode

i.e. of extended units of meaning (Sinclair 1996). GATE is optimally designed for linguistic annotation. Even though ELAN can also be used for linguistic annotation, it would not be adequate for processing multiple speakers' conversations. ELAN has been primarily developed for psycholinguistic research and each layer of annotation depends on the principal layer attributed to a single speaker. This would be impracticable in the case of our corpus. The change of tool has made impossible the direct access to the audiovisual material and necessitated the simultaneous use of both platforms.

3.3 Reference corpora

Although the context of situation provides the main clues on how/why words are used in a specific utterance and what was really meant by it, it cannot be reasonably argued that there is one and only one meaning in each utterance, or a single and precisely determined stance, nor that such meaning or stance can be fully and indisputably perceived and decoded by the analyst. This also applies to our effort to distinguish racist-oriented from neutral discourse elements. What one considers as being blatantly racist or has "labelled" as politically incorrect may be considered "neutral" by others. While well-known racial slurs, such as the lexeme *nigger*, are marked as offensive in all contemporary dictionaries, it can be observed that some of these terms used to be neutral in the past. In order to avoid – as far as possible – being influenced by preconceptions and personal beliefs, especially in view of the cross-linguistic nature of the investigation, the interpretation of how racist markers are used in discourse was based on the evidence provided by reference corpora, one for each of the three languages considered.¹²

In the comparative analysis the intuition of the annotator was supplemented with the interpretation of data from the corpora pre-loaded on the SketchEngine platform (Kilgarriff et al. 2004).¹³ These corpora are: enTenTen12 (11+ billion words), GkWaC (124 million words) and esTenTen11 (2.1 billion words), respectively for English, Greek and Spanish.¹⁴ The selected corpora offer texts collected from the Web and, as such, include a variety of text genres and a variety of lan-

¹² "Reference corpora [...] can be used as benchmarks for special corpora. Whenever we do not want to look at standard language as a whole but at some special phenomenon we happen to be interested in, we usually have to compile a corpus that fits our research focus" (Teubert & Čermáková 2007: 68). In Translation Studies, the use of reference corpora as discourse benchmarks is exemplified, *inter alia*, in Kenny (1998: 516).

¹³ www.sketchengine.co.uk.

¹⁴ For a definition of the TenTen collection of corpora, see Jakubíček et al. (2013).

guage uses (both formal and informal). Furthermore, the SketchEngine system allows the visualisation of an extended co-text of concordances. Web corpora also include authentic texts that have not been stylistically filtered and can thus present higher frequencies of informal language, slang and insulting words compared to other general reference corpora (e.g., in the case of Greek, the Corpus of Greek Texts/SEK¹⁵ and the Hellenic National Corpus/HNC¹⁶), which usually comprise only authoritative texts or texts that have previously been published in printed form or broadcast and, in that sense, have undergone pre-print editing and perhaps been subjected to cultural and linguistic “filtering” prior to their publication. In most cases, the content of such general reference corpora is “mainstream”/standard texts.¹⁷

4 Corpus annotation: Analysis and examples

4.1 Implementing the Appraisal Theory

This study focuses on the “negative” expressions of racism, i.e., on the discursive emphasis on negative and/or de-emphasis of positive “things” about *Them*, of which there are many in the corpus. It is beyond our scope to examine the emphasis on positive things and/or the de-emphasis of negative things about *Us*, although these are, more often than not, the other side of the same coin (van Dijk 2000b: 44).

For the purposes of our work we have partially adapted the Appraisal Theory model by focusing on the classification and “graduation” of attitudes, including the graduation of engagement features, which are labelled “strength” in our schema. Thus, strength is linked to:

- the valence of an evaluation and its intensification/mitigation; and
- the intensity of the speaker’s engagement with the evaluation, given that our focus is on the degree of the negative attitude expressed by the speaker as a whole.

¹⁵ SEK is provided by the University of Athens <http://sek.edu.gr>.

¹⁶ HNC is provided by the Institute for Language and Speech Processing (ILSP) <http://hnc.ilsp.gr>.

¹⁷ Teubert & Čermáková (2007: 65–66), referring to English, define its “standard” form as corresponding to the “private annual reading load of educated middle class citizens” and go on to describe a possible formulation of this definition, in corpus design terms. Such an analogy could also be made for Greek, even though it is not always clear or substantiated how the textual sub-categories would be included in such a corpus.

The attitude types used are summarised as follows:

- *Affect*: emotions and emotional reactions;
- *Judgement*: evaluation of behaviours;
- *Appreciation*: evaluation of phenomena, including aesthetics.

Our main focus is therefore on exploring the interpersonal nature, the tenor of the discourse and the social relations of the characters of the films, as far as racist stances are manifested in discourse, elements of which are the evaluation of behaviours and situations and the expression of negative emotions and opinions (see Fotopoulou et al. 2009). In this sense, we have categorised and manually annotated “[a]ttitudes [...] divided into three regions of feeling, ‘affect’, ‘appreciation’ and ‘judgement’” (Martin & White 2005: 35–43).

For the purposes of our analysis, we annotated every segment of racist attitude of the film characters (appraisers) that express negative evaluations towards a person or a group of another ethnicity/“race”/religion (appraised entity). All candidate instances are interpreted and classified as instantiations of a type of attitude towards “Others”. In this sense, the utterance *Minorities don’t give two shits about this country* (American History X), taken semantically, simply informs us about the indifference (affect) of minorities towards *this country*. However, considering how it is used in the given context, we focus on the interlocutor’s intended meaning, which is to implicitly criticise the members of minorities as indifferent free-loaders who just want to exploit the country. Therefore, it is annotated as an implicit judgement.

4.2 Attitude types and usage perspective

A systematic classification of racist attitudes requires the clearest possible definition for each category, since the boundaries among attitude types are not always accurate and undoubted. This is not surprising, given that “[i]n a general sense, affect, judgement, and appreciation all encode feeling” (Martin 2000: 147). It cannot be argued that negative judgements or appreciations bear no traces/nuances of affect and that speakers can either express emotions or judge. As a matter of fact, “[a]ffect can perhaps be taken as the basic system [...]” (ibid.), as an immanent or emerging characteristic of every attitude. Through this perspective, “[a]s judgement, affect is re-contextualised as an evaluation matrix for behaviour [...] [a]s appreciation, affect is re-contextualised as an evaluation matrix for the products of behaviour (and wonders of nature)” (Martin 2000: 147).

In this sense, we have annotated as instances of *affect* all expressions of emotions, as signs of emotional reactions of the characters related to the specific discourse type. Further, *appreciation* was defined as evaluations, mainly aesthetic, of humans as entities, as evaluations not related to their behaviour but to their characteristics such as colour, ethnicity, religion, physical characteristics and supposed inherent characteristics. Such assessments are negative markers of difference. Finally, *judgement* which “deals with attitudes towards behaviour, which we admire or criticise, praise or condemn” (Martin & White 2005: 42) applies to cases that can be defined as rationalisations of a fact or as reflecting a causal relation between things or facts. On the contrary, in *appreciation*, the speaker’s stance is intuitive or dogmatic, i.e. non-refutable in the specific context of situation. In our opinion, this formalisation defines more objectively the classification of attitudes compared to the broader and more inclusive definition offered by Martin & White (2005: 42–45).

The basic tri-fold categorisation of attitude as applied in this study is as follows:

- *Affect*: characterises negative emotions and emotional reactions towards Others, principally instances that are indicative of hate and anger based on or evoked by racial/ethnic/religious differences.
- *Appreciation*: characterises negative evaluations of Others based on their inherent characteristics, or on characteristics that are presented as such, dogmatically used as sufficient reasons for negative evaluations.
- *Judgement*: characterises negative evaluations of the behaviour of Others, judged as people that act in relation to their racial/ethnic/religious difference.

In the paragraphs below, these types are exemplified.

4.2.1 Affect

- (1) Your mother, she *hated them niggers* too (*Monster’s Ball*)
- (2) That means *not welcome* (*American History X*) [utterance addressed to a Jew while the speaker reveals his swastika tattoo]
- (3) Take your fucking pizza piece and *go the fuck back to Africa* (*Do the Right Thing*)
- (4) Yeah, *fuck off, you Paki bastards* (*This is England*)
- (5) *What the hell are those niggers* doing out there? (*Monster’s Ball*)

4.2.2 Appreciation

- (6) We'll let the *niggers, kikes and spics* grab for their piece of the pie (*American History X*)
- (7) *A bunch of people who aren't even citizens of this country [...]* (*American History X*)
- (8) She *smells like fish and chips and guacamole* (*American History X*)
- (9) Three and a half million of us, who can't find fucking work because *they're taking them all, because it's fucking cheap labour* (*American History X*)
- (10) How come *niggers are so stupid?* (*Do the Right Thing*)
- (11) Magic, Eddie, Prince, are not *niggers*. I mean, they're not *black*, I mean... (*Do the Right Thing*)

4.2.3 Judgement

- (12) Look at these little *fucking sewer rats* (*This is England*) [referring to young immigrants playing in a yard]
- (13) *Immigration, AIDS, welfare, those are problems of the black community, the Hispanic community, the Asian community* (*American History X*)
- (14) One in every three black males is in some phase of the correctional system. Is that a coincidence or do these people *have like a racial commitment to crime?* (*American History X*)
- (15) All right. Well, you know what I can't do? I can't look at you without thinking about the *five or six more qualified white men who didn't get your job* (*Crash*)
- (16) He's one of those *proud to be nigger people* (*American History X*)

Although “interpersonal epithets” (see Halliday & Matthiessen 2014: 376–377), e.g. evaluative adjectives, are the most obvious evaluative device of language, the lexico-grammatical choices that express attitude are infinite, especially if we consider that evaluative uses of language can be present in discourse both explicitly and implicitly (see Munday 2012: 23). As shown in the above examples, the

lexico-grammatical means to express attitude are vast and not limited to closed semantic or grammatical categories.

In addition, phenomena investigated from another point of view in previous studies are also evident in our data, but are further analysed as instantiations of evaluative attitudes. Referential/nomination strategies and predication strategies, such as racial slurs and metaphors (*sewer rats*) analysed by Reisigl & Wodak (2001) as a means to categorise membership, are analysed here as evidences of the speaker's stance. Thus, we observe the presence of racial slurs, such as *nigger*, in all three types of attitude:

- used to express mere anger and hate, as in (1) and (5),
- used in appreciations just to refer to Others in a disparaging manner, presenting the fact of belonging to other racial groups and/or having their “inherent” characteristics as being *per se* negative, as in (6) and (11), or
- used in judgements to negatively evaluate a certain person's behaviour, which is presented as related to the fact that he/she is black (12).

The interpretation of each instance is based on the context of situation, and is further based on the presence of various markers, either explicitly stated, as in (1) where the verb *hate* is used, or using interjections such as *go the fuckback*, *fuck off* and *what the hell* [in (3), (4), and (5)].

On the other hand, in (6) racial slurs are used instead of “neutral” ethnic denominations as disparaging terms, simply to mark the inferiority of the mentioned groups. The utterance in example (11) is a response to the interlocutor's argument that, despite his constant negative attitude towards black people, all his favourite celebrities (Eddie Murphy, Magic Johnson and Prince) are black. It is a representative example of how a highly marked racial slur is used as a negative appreciation, to evaluate people as being “nice” or “bad”.

Accordingly, our data includes many convictions and stereotyped visions of Others (7) and recurrent topics or *topoi*, as “common-sense reasoning [that is] typical for specific issues” (van Dijk 2000c in Baker et al. 2008: 299, note 21; see also Reisigl & Wodak 2001: 74–76). Examples are the *topos* of finance (9), the *topos* of threat (13, 14) and the *topos* of justice or equal opportunities (15). Such visions can be used as appreciations or judgements, i.e. presented as either negative phenomena, as in (9), or as criticism of the behaviour of Others, as in (13), (14) and (15).

4.2.4 Type overlaps

As mentioned already, it is normal that the boundaries between categories are not always clear-cut: thus, in cases that could belong to more than one category, we have used double annotation: this is both methodologically permissible and of course technically possible. This allows for a subsequent analysis on both levels, e.g. of *affect* and *appreciation*, for the sake of contrastive analysis within the two categories, and hence for further refining the classification/annotation scheme.

- (17) *You gold-teeth, gold-chain-wearing, fried-chicken and biscuit-eating monkey, ape, baboon, big-thigh, fast running, high-jumping, spear chucking, 360-degree basketball-dunking, titsoon, spade, moulignon!*

In this sense, according to our definition of attitudes, the utterance in (17) used by an Italian-American in an aggressive manner to express hatred directed to a black person, shows negative affect. At the same time, the long list of epithets enumerated represent appreciations referring to his interlocutor.

4.3 Attitude features

Attitudes are also analysed in terms of their features.

4.3.1 Implicit and explicit attitudes

As mentioned above, we always categorise attitudes expressed towards persons or groups of other ethnicities/“races”/religions. In many cases, attitudes are not explicitly stated in the text, but evoked, expressed implicitly. Such an interpretation can rely on common knowledge, on the co-text and on the context of the situation.

“If by expressing meaning A, language users (also) mean B, such an implication can be reconstructed by recipients only on the basis of inferences from culturally shared knowledge of language meanings (e.g. as represented in the lexicon of the language) or more generally on the basis of shared knowledge, including particular knowledge about the knowledge of the speaker” (van Dijk 1995: 168).

Thus, in example (8) above, one should know that *fish and chips and guacamole* refer to the culinary traditions of Latin Americans in order to properly interpret

it, whereas in example (12) knowledge about the policies of affirmative action against racial discrimination in the USA is crucial in order to interpret the utterance correctly. Moreover, in example (12) the term *sewer rat* is used metaphorically to designate useless people that cause problems to society.

4.3.2 Irony

Irony is also a case of implicit meaning, a pragmatic phenomenon used to express a meaning contrary or different to the literal one. Once again, its recognition depends on the situational context which indicates something different than the apparent meaning of the utterance, the speaker's tone of voice or the "interpreters' assumptions about the beliefs or values of the text producer" (Fairclough 1992: 123). For instance, the utterance in example 18 is used to depreciate African-American literature:

(18) What is it, Black History Month? (*American History X*)

4.3.3 Indirect/Direct attitudes

Some utterances do not express a racist attitude but are still indicative of the social roles of the participants and can reflect racism as internalised.¹⁸ These are cases where speakers comment the racist stance of others and are annotated as indirect references to racist attitudes.

(19) Man's singing about lynching niggers. "Gonna buy me a rope and lynch me a nigger" (*Crash*)

(20) Your partner's a racist prick (*Crash*)

4.3.4 Polarity

Polarity refers to the positive or negative dimension of an attitude, distinguishing positive from negative affect, appreciation or judgement. As mentioned already, for the aims of our study we focus on negative instances.

¹⁸ *Internalised racism* is defined as the situation in which individuals, groups and cultures that have been subjected to racism and oppression, shift this racism to oppress themselves and others who have experienced racism and discrimination (Lawrence 2002: 92).

4.3.5 Strength

The strength of each instance is taken into account and an indication of low, medium or high valence is given to each instance. Admittedly, this is the most subjective parameter of our effort, so that in order to decide on the strength of an utterance inter-subjective agreement among the annotators was required in most cases. The collocational behaviour of the lexemes examined was assessed with the help of the English reference corpus.

4.4 Annotation scheme overview

The attitude classification outlined above and used as annotation scheme in the reported project can be schematised in Figure 4.

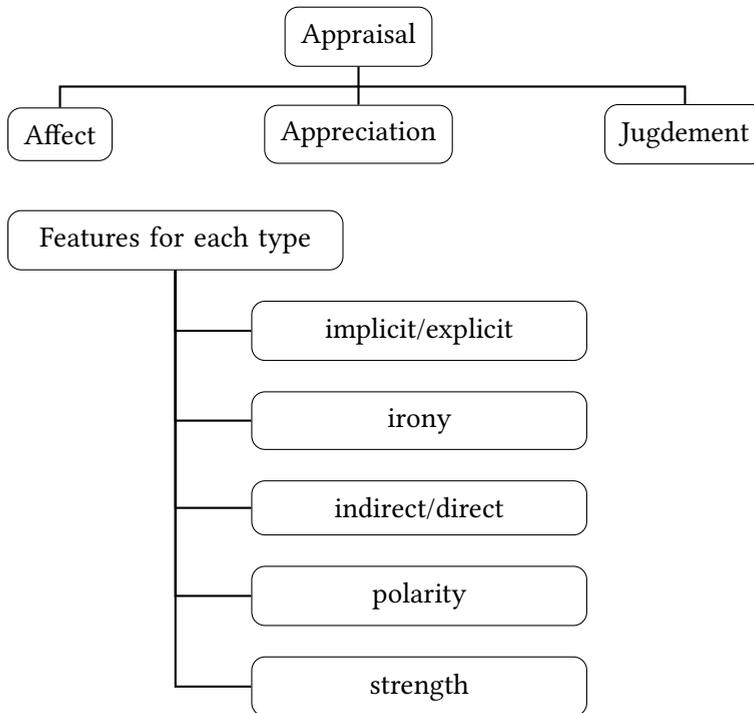


Figure 4: Corpus annotation scheme

5 Register shifts: Analysis and examples

This study investigates how racist/heterophobic discourse is transferred when a film is translated into socio-cognitively distinct and somewhat remote linguistic systems. It tries to systematise changes of tenor (Halliday 1978) in translation by means of a comprehensive classification of attitudes and their cross-linguistic mapping.

We therefore believe that retrodictively (von Wright 1971; Chesterman 2008) the diachronic study of discourse features, and more generally of racism, in translation could also benefit from such a hermeneutic approach. In the examples below, the shifts observed in discourse transfer have been explained in the light of the relationship between participants in discourse.

5.1 Examples

(Examples 21–24 are sourced from *American History X*; examples 25–27 are sourced from *Crash*.)

- (21) [EN] And now some fucking Korean owns it who fired *these guys* and is making a killing because he hired *40 fucking border jumpers*
 [EL] Τώρα το 'χει ένας Κορεάτης, που απέλυσε τους δικούς μας και θησαυρίζει επειδή προσέλαβε *λαθρομετανάστες*
 [Back translation] Now a Korean owns it who fired *our guys* and is making a killing because he hired *illegal immigrants*

Derek, a young skinhead, gives a speech to the rest of the gang members, trying to convince them to attack a supermarket owned by immigrants. Throughout his speech, negative attitudes towards immigrants are abundant. Among other arguments, he uses a variety of *topoi* as in, e.g. *immigrants take our jobs*. He uses rather colloquial expressions and the register of his speech is highly informal, indicative of the brotherhood relations among the in-groups. In terms of Appraisal Theory, this utterance is considered to be a highly marked negative judgement about immigrants.

If we concentrate on the last part of the utterance, that is *40 fucking border jumpers* and its respective Greek version rendered as *λαθρομετανάστες*, we notice immediately that the strength of the judgement is significantly altered. A closer look at each component of the TL unit of meaning reveals that *border jumper*, an apparently neutral term describing an action, is used as a depreciative, non-fixed, and possibly colloquial, synecdoche of *immigrants of Hispanic/Mexican origin*. An analysis of the term in enTenTen12 has returned only 57 hits of *border*

jumper(s), i.e. a negligible frequency in the enTenTen12 corpus (0.0 per million). Furthermore, as an analysis of the concordance lines reveals (see Figure 5, below), the term is used almost exclusively in a negative and highly disparaging sense (*border jumpers want our wealth; drug smugglers, human traffickers, border jumpers and other assorted criminals; border jumpers are slapping those legal immigrants*). Moreover, the presence of *fucking* (in the cluster *fucking border jumpers*), as an intensifier, as well as the emphatic mention to the number of immigrants employed, reinforce the overall negative prosody of the judgement, making it highly negative. On the other hand, the Greek translation of the utterance is limited to *λαθρομετανάστες* [clandestine immigrants], which is a generic term with no real connotation about the specific origin of the immigrants in question. By contrast, in GkWaC, the term *λαθρομετανάστης* has a frequency of 5.9 per million and its “ideologically neutral” synonym *παράνομος μετανάστης*¹⁹ appears with a frequency of 0.7 per million. As to its usage profile, the term in question appears in various text genres and, most importantly, belongs to “standard” Greek as it is present in authoritative language.²⁰ The rendition of the utterance in Greek is a translation shift, in both field and tenor. The judgment loses its strength and maintains only the negative nuance inferred by the context of situation, as well as by the invented contrastive relation (Fairclough 2003: 87–89), i.e. by the contrast between *δικούς μας* ‘our guys’ and *λαθρομετανάστες* ‘illegal immigrants’, that has been added explicitly in the TT.

- (22) [EN] I mean, Christ, Lincoln *freed the slaves*, what, like hundred and thirty years ago. *How long does it take to get your act together?*
[EL] Ο Λίνκολν *απελευθέρωσε τους σκλάβους* πριν 130 χρόνια. *Πόσον καιρό χρειάζεσαι για να γίνεις άνθρωπος;*
[Back translation] Lincoln *freed the slaves* a hundred and thirty years ago. *How long does it take to become a human?*

During a family dinner, Derek argues with his professor (who apparently has an affair with his mother) about riots in the black neighbourhoods. Example (22)

¹⁹ The use of the prefix “λαθρο-” (a derivative of the adjective *λαθραίος* clandestine, smuggled) to designate economic or political immigrants has been criticised by human rights and political organisations as being negatively loaded, even though this is not always the case, since language economy, not surprisingly, seems to opt for the single-word designator (*λαθρομετανάστης*) rather than for the presumably more neutral two-word unit (*παράνομος μετανάστης*). This is apparent also in the concordances derived from GkWaC, where most occurrences do not have a negative connotation. The neutral (and hence stabilised as politically correct) designator *παράνομος* (illegal) is used instead by the administration.

²⁰ See above, note 17.

Query **border, jumper** 57 (0.0 per million)

Page 1 of 3 [Next](#) | [Last](#)

doc#109056	since there are other ways of identifying border jumpers , such as “the kind of dress ... right down
doc#171898	identifiable (Muslims in Russia, Hispanic border jumpers in the US), both are unified (at least
doc#625986	my mind. <i></p><p></i> Profiling? Well, when the border jumpers are coming in from Mexico it shouldn't
doc#629656	because of the criminal prosecution of repeat border jumpers . <i></p><p></i> ... “It's unclear ... [whether]
doc#666369	himself important. That must remind you of border jumpers . <i></p><p></i> AJ <i></p><p></i> Libertarian <i></p><p></i> Le:
doc#1011415	in the common law status of “outlawry.” Border jumpers and immigrants who want our wealth and
doc#1337820	taking jobs. We need to quit focusing on border jumpers and look at the real problem. As stated
doc#2219845	had enough of nobody doing anything about border jumpers : Source: Los Angeles to boycott Arizona
doc#3395787	known fence breach points” as a signal to border jumpers . <i></p><p></i> Other seemingly innocent activitie
doc#3442825	shout. Not wishing to be hauled in as a border jumper , Holmdahl leaped into the frigid Rio Grande
doc#3498887	mexico sends its millions of illegal alien border jumpers to invade the U.S., cry “racism”, pose
doc#3658404	forementioned drug smugglers, human traffickers border jumpers and other assorted criminals. I wonder
doc#5057516	identified that NSW had the highest number of border jumpers in 2002-2003. In total, 125,195 people
doc#5206024	immigration reforms go through and stopping the border jumpers . Rather than fill up a country that is
doc#5607550	prohibited from enforcing the law to stop these border jumpers , your families in Plainsville, Ohio, or
doc#5811734	worked very hard to get in here legally. Border jumpers are slapping those legal immigrants in
doc#5987724	the transit situation in which would-be border jumpers are retained blurs the distinction between
doc#6113973	borders are all but non-existent, and if a border jumper happens to stake their claim in this country
doc#6375465	target undocumented illegal immigrants and border jumpers . <i></p><p></i> Government officials from South
doc#6788903	border? Where do all the illegal Zimbabwean border jumpers go to find jobs in South Africa? Why do

Figure 5: Concordance lines, *border jumpers*, GkWaC, in SketchEngine

comes to reinforce his argument that the sheer number of jailed black people proves their racial commitment to crime. The idiom used in this example, *get one's act together*, has the meaning of getting organised and being on schedule. The utterance is implicitly ironic. In our study, the source segment has been noted as a negative judgement of medium strength, this being mainly due to the ironic nature of both the argument and the reference to the end of slavery. On the other hand, the Greek translation uses the idiom *γίνομαι άνθρωπος* ‘become human’ (GkWaC frequency: 1.9 per million) which has the meaning of “becoming an ethical and useful citizen”, thus implying a shift towards a stronger negative judgement about the social attitude of black people.

- (23) [EN] We'll let the *niggers, kikes and spics* grab for their piece of the pie.
 [ES] Dejemos que *negros y latinos* se lleven su parte.
 [Back translation] Let the *blacks* and the *Latinos* get their part.

The leader of the skinhead gang, a middle-aged male, tries to convince Derek that their organisation is going to stop being a small gang and will now grow into something very serious and powerful all over the country. He explains how he plans to act in order to achieve it.

In example (23) above, we mark the use of three racial slurs, *nigger* (enTenTen12 frequency: 0.9 per million) as highly negative appreciations towards Jews, Latinos, and black people respectively. In the Spanish TT, the racial slurs of the original are shifted towards neutralisation (*negros* and *latinos*), while the reference to Jews (*kikes*) is eliminated. In all, the negative appreciations that are inherent in racial slurs are eliminated. Although in Spanish there is a lexeme, *negrata*, which has the same derogatory connotation as *nigger*, translating *nigger* as *negro* is a recurrent practice in our corpus. However, *negrata* is a term with a very low frequency (only 97 tokens in esTenTen11), while *nigger* appears with a frequency of 0.9 per million in enTenTen12. This observation could be indicative of the reasons that made Spanish subtitlers opt for the neutral term, avoiding the use of an uncommon term. On the other hand, *spic*, a term that could have been translated as *sudaca* (a Spanish racial slur with presumably similar connotations with 431 tokens in esTenTen11) is also avoided. In both cases, the translator opts for neutralising the rendition of the original utterance.

- (24) [EN] Name your price, *cracker*.
[ES] Di tu precio, *blanco*.
[Back translation] Name your price, white guy.
[EL] Πες το ποσό, *βλάχο*.
[Back translation] Name your price, country bumpkin.

During a basketball game in the neighbourhood court, members of the skin-head gang start to quarrel with members of a “black gang”. Derek has a bet; he proposes a “whites against blacks” game. The answer in (24) comes from one of the opponents, indicating that they accept the bet. In example (24) above, the disparaging term *cracker* (showing negative appreciation, i.e. for a poor white person, usually from the South) is translated as *blanco* ‘white’ in Spanish, but as *βλάχο* ‘country bumpkin’ in Greek. In this case, the Spanish subtitler succeeds in maintaining the racial nuance of the term, although the strength of the negative appreciation is diminished, while the Greek subtitler eliminates the racial reference and only transfers the aggressive tone of the dialogue.

- (25) [EN] - Do you speak English? - I am speaking English, you stupid cow!
[EL] - Θα μιλήσετε Αγγλικά; - Μιλάω Αγγλικά!
[Back translation] - Are you going to speak English? - I speak English!

In example (25) above, an Asian woman enters a hospital screaming the name of her husband. The intuitive reaction of the nurse is to ask her if she speaks English, a reaction reflecting the stereotype that immigrants do not speak the language of the host country. Thus, we mark this instance as a low strength negative appreciation, yet expressed in a polite manner. The Greek subtitler opts for a more aggressive-impolite way in rendering this question, *θα μιλήσετε Αγγλικά* ‘Are you going to speak English?’ strengthening the negative valence of the utterance. There are many similar examples in our corpus, pointing to the linguistic identity as a marker of difference. Sella-Mazi (2001: 111) argues that language is involved in matters of political and social texture, by functioning as the defining element of the nature of multiple human groupings, either positively by delineating “Us”, or negatively, by excluding allophones from the said group: in this case, the interlocutors are self-determined contrastingly, both within and outside a linguistic group (the “Others”).

- (26) [EN] Stupid *wetback*
 [ES] Estúpida *sin papeles*
 [Back translation] Stupid undocumented immigrant

In (26), an Asian and a Latin American woman are involved in a car crash. While they quarrel, the first one calls the other a *wetback*. The term is highly disparaging and refers to illegal Latin Americans (especially Mexicans) as a descriptor of the way Mexicans enter the US by crossing the Rio Grande. Therefore, it has been marked as a negative judgement utterance. Although the term is decades-old and has been used even in the title of a deportation programme of the US in 1954 (the so-called “Operation Wetback”, see Hernandez 2006), today it is used in a highly derogatory manner as a racial slur. As shown in Figure 6, the most significant collocations of *wetback* (sorted by Mutual-Information, in a query window of -10 to +10 tokens) are other racial slurs, especially in their context of usage (e.g. *mojado*, *beaners*, *spics*, *kike*, *nigger*, *chink*, *greaser*, *lowlife*, etc.).

On the other hand, the Spanish translation uses *sin papeles* ‘undocumented immigrants’. This is a rather neutral term to refer to illegal immigrants. In esTenTen11, the most significant collocates of *sin papeles* (see Figure 7) are emotionally neutral (*inmigrante* ‘immigrants’, *empadronar/empadronamiento* ‘inclusion in the town registry’, *redadas* ‘raids’, *patera/pateras* ‘dinghy’) and an analysis of the concordances shows that the term is used also in texts, in support of the human rights of immigrants.

Collocation candidates

Page [Next >](#)

	Freq	MI
P N mojado	5	21.548
P N Wetback	6	19.708
P N wetback	28	19.418
P N beaners	6	19.200
P N spics	6	18.866
P N kike	7	17.176
P N Belling	6	17.057
P N Chink	3	17.027
P N nigger	68	16.606
P N spic	5	16.273
P N greaser	3	15.815
P N lowlife	5	15.319
P N OPERATION	3	15.083
P N coon	6	14.741
P N Boilers	3	14.629
P N chink	7	14.472
P N carload	3	14.448
P N whitey	3	14.371
P N Operation	73	14.253

Figure 6: Collocates of *wetback* in enTenTen12, sorted by Mutual Information (MI) in SketchEngine

Collocation candidates

Page [Next >](#)

	Freq	logDice
P N empadronar	33	8.094
P N inmigrantes	724	8.038
P N inmigrante	122	7.486
P N redadas	21	7.219
P N empadronamiento	30	7.018
P N pateras	21	6.908
P N patera	21	6.731
P N emigrantes	53	6.714
P N Clandestino	12	6.658
P N Vic	28	6.633
P N regularización	26	6.451
P N expulsiones	16	6.362
P N deportación	12	6.342
P N regularizar	16	6.291
P N repatriados	10	6.221
P N emigrante	15	6.148

Figure 7: Collocates of *sin papeles* in esTenTen11, sorted by logDice in SketchEngine

In other words, the racist tone of the insult, even though it is still present in the Spanish text, is reduced in the interpersonal component of the utterance.

- (27) [EN] - You wanna buy these Chinamen? - Don't be ignorant. They're Thai or Cambodian. Entirely different kind of *chinks*.
 [ES] - ¿Vas a comprar esos chinos? - No seas ignorante. Serán tailandeses o camboyanos. Son unos *amarillos* distintos.
 [Back translation] Should be Thai or Cambodian. They are different yellow people.

Once again, a character of *Crash* uses a racial slur to refer to a group of Asians. He, quite ironically, has just decided to set them free instead of accepting money to “sell” them with the truck he robbed and proved to have been used for human trafficking. This is a negative appreciation texteme. Once again, *chink* (queried in SketchEngine as a noun) was found to collocate with other racial slurs in our reference corpus. However, a similar query for *amarillo* ‘yellow’ does not yield any results, since as a noun it refers also to the colour, and in the tool used, semantic disambiguation is not possible. However, the use of a colour to designate a race does not necessarily indicate a racist stance, given that the utterance is not directed at Asians. This explanation is consistent with the definition of *amarillo*, taken from the online version of the *Diccionario de la Real Academia Española (DRAE)*, which is not marked as derogatory, but as simply referring to the Asian race.²¹

5.2 Summary of findings

Each language/culture produces and stabilises forms of expressing (racist) meanings that are unknown or at least asymmetrically²² represented in other languages and cultures. Racial slurs are a clear example of this and constitute crucial points for the translator, who often mitigates or omits them. Even though the mitigation of racial slurs is the general tendency in our corpus, there are also cases of over-toning of racist attitudes, e.g. in examples (22) and (25); in both these examples, the (racist-oriented) interpersonal meaning is intensified in the TL, though both target utterances lack marked epithets (slurs). Such instances should be explored further. This article has discussed different types of register

²¹ “Dicho de un individuo o de la raza a que pertenece: De piel amarillenta y ojos oblicuos. Apl. a pers., u.t.c.s.” <http://goo.gl/oelvVZ>.

²² For a discussion on cultural asymmetry in translation, a concept coined in TS by Even-Zohar (2005), see e.g. Klaudy et al. (2012).

shifts, without attempting to provide a systematic analysis (for instance, we did not consider the many instances of racist discourse which did not undergo significant register shift in translation). Such a systematic analysis, which shall be pursued in further research, will hopefully provide a more “sustainable” picture of how racist discourse is handled in translation.

6 Conclusions

This paper has presented a research, based on the PhD thesis of the first author, aimed at investigating the translation of racist discourse in Anglophone films subtitled in Greek and Spanish.

To this end, we have developed a linguistic annotation model in order to systematically categorise racism-related utterances in original films and in their subtitled versions. Instances of stereotyped views, prejudices, racist attitudes and emotions triggered by racism were coded using an annotation scheme based on Appraisal Theory.

The reference corpora used for the analysis were extremely useful, though with some limitations. Firstly, they were neither corpora of spoken discourse nor balanced corpora. Secondly, some of the terms looked up have very low frequencies and therefore do not allow for a safe description. Thirdly, it was not possible to find reliable evidence for ambiguous terms such as *amarillo* or *sin papeles*. In many cases the meaning of a racist expression could be interpreted only by analysing its context in the film.

Our analysis of register shifts in translation, based on a Systemic Functional Linguistic approach, is promising for the descriptive study of the socio-culturally marked discourse of racism and aims to serve as an explanatory basis for addressing broader questions:

- Is it possible, and if so how, to refine the definitions of heterophobia that have formed part of our initial motivation in functional linguistic terms?
- Which is the relation between cinema and socio-linguistic reality in the perception of xenophobia?
- What are the implications of such an analysis, in relation to the comprehension of racist discourse and its root causes in the modern Greek and European linguistic and cultural reality?

Last but not least, our findings so far point to the assumption that such an approach could, indeed, be linked systematically to the critical study of the role

of translation in the diachronic development of the sociolinguistic dimension of racism.

References

- Asher, Nicholas, Farah Benamara & Yvette Yannick Mathieu. 2009. Appraisal of opinion expressions in discourse. *Linguisticae Investigationes* 32(2). 279–292.
- Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michał Krzyżanowski, Tony McEnery & Ruth Wodak. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society* 19(3). 273–306.
- Batsalia, Frideriki & Eleni Sella-Mazi. 2010. *Linguistic approach to the theory and didactics of translation*. 2nd edition. Athens: Papazisis.
- Brugman, Hennie & Albert Russell. 2004. Annotating multimedia/multi-modal resources with ELAN. In *Proceedings of the forth international conference on language resources and evaluation*. Paris: ELRA.
- Chesterman, Andrew. 2008. On explanation. In Anthony Pym, Miriam Shlesinger & Daniel Simeoni (eds.), *Beyond descriptive translation studies*, 363–379. Amsterdam: John Benjamins.
- Cunningham, Hamish, Diana Maynard, Kalina Bontcheva & Valentin Tablan. 2002. Gate: A framework and graphical development environment for robust NLP tools and applications. In *Proc. 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*. Philadelphia.
- Eggins, Suzanne & Diana Slade. 1997. *Analysing casual conversation*. London: Cassell.
- Essed, Philomena. 1991. *Understanding everyday racism: An interdisciplinary theory*. London: Sage.
- Even-Zohar, Itamar. 2005. Laws of cultural interference. In *Papers in culture research*. <http://goo.gl/Fvw4ZM>.
- Fairclough, Norman. 1985. Critical and descriptive goals in discourse analysis. *Journal of Pragmatics* 9. 739–763.
- Fairclough, Norman. 1992. *Discourse and social change*. Cambridge: Polity Press.
- Fairclough, Norman. 2003. *Analysing discourse. Textual analysis for social research*. London: Routledge.
- Forster, Marc. 2001. *Monster's ball*. Corpus. US: Lions Gate Films.

- Fotopoulou, Aggeliki, Marianna Mini, Mavina Pantazara & Argiro Moustaki. 2009. La combinatoire lexicale des noms de sentiments en grec moderne. In Iva Navacova & Agnès Tutin (eds.), *Le lexique des émotions*. Grenoble: ELLUG.
- Giddens, Anthony. 2009. *Sociology*. 6th edition. Cambridge: Polity Press.
- Gottlieb, Henrik. 1997. *Subtitles, translation & idioms*. Copenhagen: University of Copenhagen.
- Gottlieb, Henrik. 2005. Multidimensional translation: Semantics turned semiotics. In *Proceedings EU High Level Scientific Conference Series: Multidimensional Translation (MuTra)*. Copenhagen.
- Haggis, Paul. 2004. *Crash*. Corpus. US: Lions Gate Films.
- Halliday, M. A. K. 1978. *Language as social semiotic*. London: Arnold.
- Halliday, Michael Alexander Kirkwood & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael Alexander Kirkwood & Christian Matthiessen. 2014. *Halliday's introduction to functional grammar*. 4th edition. London: Routledge.
- Hatim, Basil & Ian Mason. 1997. *The translator as communicator*. London: Routledge.
- Hernandez, Kelly Lytle. 2006. The crimes and consequences of illegal immigration: a cross-border examination of Operation Wetback, 1943-1954. *Western Historical Quarterly* 37. 421-444.
- Jaffe, Alexandra. 2009. *Stance. Sociolinguistic perspectives*. Oxford: OUP.
- Jakobson, Roman. 2012. On linguistic aspects of translation. In Lawrence Venuti (ed.), *The Translation Studies reader*, Third edition. London: Routledge.
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen corpus family. In Andrew Hardie & Robbie Love (eds.), *Corpus linguistics 2013 abstract book*. Lancaster: UCREL.
- Johansson, Stig. 2003. Reflections on corpora and their uses in cross-linguistic research. In Federico Zanettin, Silvia Bernardini & Domenic Stewart (eds.), *Corpora in translator education*. Manchester: St. Jerome Publishing.
- Kaye, Tony. 1998. *American History X*. New Line Cinema.
- Kenny, Dorothy. 1998. Creatures of habit? What translators usually do with words. *Meta* 43(4). 515-523.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The SketchEngine. In *Proc EURALEX 2004*, 105-116. Lorient, France.
- Klaudy, Kinga, Hannu Kemppanen, Marja Jänis & Alexandra Belikova. 2012. Linguistic and cultural asymmetry in translation from and into minor languages. In Hannu Kemppanen, Marja Jänis & Alexandra Belikova (eds.), *Domestication and foreignization in Translation Studies*, 33-48. Berlin: Frank & Timme.

- Kress, Gunther. 2010. *Multimodality: a social semiotic approach to contemporary communication*. London: Routledge.
- Lawrence, Duncan. 2002. Racial and cultural issues in counselling training. In Aisha Dupont-Joshua (ed.), *Working inter-culturally in counselling settings*, 88–105. London: Routledge.
- Lee, Spike. 1989. *Do the right thing*. Corpus. US: Universal/40 Acres & a Mule Filmworks.
- Martin, James Robert. 2000. Beyond exchange: Appraisal systems in English. In Susan Hunston & Geoffrey Thompson (eds.), *Evaluation in text*, 142–175. Oxford: OUP.
- Martin, James Robert & Peter R. R. White. 2005. *The language of evaluation: Appraisal in English*. London: Palgrave Macmillan.
- Mason, Ian. 2001. Coherence in subtitling: The negotiation of face. In Frederic Chaume & Rosa Agost (eds.), *La traducción en los medios audiovisuales*, 19–31. Castellon: Jaume I University of Castellon.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: CUP.
- Meadows, Shane. 2006. *This is England*. Corpus. UK: Optimum Home Releasing.
- Memmi, Albert. 2000. *Racism*. Minneapolis: University of Minnesota Press.
- Mouka, Effie. 2014. Investigating translational norms in socioculturally marked cinematographic discourse. In Jenny Brumme & Sandra Falbe (eds.), *The spoken language in a multimodal context. Description, teaching, translation*, 213–228. Berlin: Frank & Timme.
- Mouka, Effie, Voula Giouli, Aggeliki Fotopoulou & Ioannis E. Saridakis. 2012. Opinion and emotion in movies: A modular perspective to annotation. In *Proceedings of the eighth international conference on language resources and evaluation*, 104–109. Paris: ELRA.
- Mubenga, Kajingulu Somwe. 2009. Towards a multimodal pragmatic analysis of film discourse in audiovisual translation. *Meta* 54(3). 466–484.
- Munday, Jeremy. 2012. *Evaluation in translation: critical points of translator decision-making*. London: Routledge.
- Pedersen, Jan. 2011. *Subtitling norms for television: an exploration focusing on extralinguistic cultural references*. Amsterdam: John Benjamins.
- Pettit, Zoë. 2005. Translating register, style and tone in dubbing and subtitling. *Journal of Specialised Translation* 4. 49–65.
- Reisigl, Martin & Ruth Wodak. 2001. *Discourse and discrimination. Rhetorics of racism and antisemitism*. London: Routledge.

- Saridakis, Ioannis E. 2010. *Text corpora and translation: Theory and applications*. Athens: Papazisis.
- Schmidt, Thomas. 2011. A TEI-based approach to standardising spoken language transcription. *Journal of the Text Encoding Initiative*.
- Sella-Mazi, Eleni. 2001. *Diglossia and society. The greek reality*. Athens: Proskinio.
- Sinclair, John. 1996. The search for units of meaning. *Textus* 9(1). 75–106.
- Taboada, Maite & Jack Grieve. 2004. Analyzing appraisal automatically. In *AAAI spring symposium on exploring attitude and affect in text*.
- Teubert, Wolfgang & Anna Čermáková. 2007. *Corpus linguistics. A short introduction*. London: Continuum.
- Toury, Gideon. 2012. *Descriptive translation studies and beyond*. 2nd edition. Amsterdam: John Benjamins.
- van Dijk, Teun Adrianus. 1993. *Elite discourse and racism*. Newbury Park: Sage Publications.
- van Dijk, Teun Adrianus. 1995. Discourse semantics and ideology. *Discourse and Society* 6(2). 243–289.
- van Dijk, Teun Adrianus. 2000a. Ideologies, racism, discourse: Debates on immigration and ethnic issues. In Jessika Ter Wal & Maykel Verkuyten (eds.), *Comparative perspectives on racism*, 91–116. Aldershot: Ashgate.
- van Dijk, Teun Adrianus. 2000b. *Ideology and discourse: A multidisciplinary introduction*. Barcelona: Pompeu Fabra University. <http://goo.gl/LmpjoQ>.
- van Dijk, Teun Adrianus. 2000c. Parliamentary debates. In Ruth Wodak & Teun Adrianus van Dijk (eds.), *Racism at the top: Parliamentary discourses on ethnic issues*. Klagenfurt: Drava.
- van Dijk, Teun Adrianus. 2002. Discourse and racism. In David Goldberg & John Solomos (eds.), *The Blackwell companion to racial and ethnic studies*, 145–159. Oxford: Blackwell.
- van Dijk, Teun Adrianus. 2004. Racist discourse. In Ellis Cashmore (ed.), *Routledge encyclopedia of race and ethnic studies*, 351–355. London: Routledge.
- Vasileva, K. 2011. 6.5% of the EU population are foreigners and 9.4% are born abroad. Population and social conditions. *Eurostat statistics in focus* 34. <http://goo.gl/SZLQ2Y>.
- von Wright, Georg Henrik. 1971. *Explanation and understanding*. Ithaca: Cornell University Press.
- Whitelaw, Casey, Navendu Garg & Shlomo Argamon. 2005. Using appraisal taxonomies for sentiment analysis. In *Proceedings of the 14th acm international conference on information and knowledge management*. <http://goo.gl/qbNIJB>.

- Zabalbeascoa, Patrick. 2008. The nature of the audiovisual text and its parameters. In Jorge Díaz-Cintas (ed.), *The didactics of Audiovisual Translation*, 21–38. Amsterdam: John Benjamins.
- Zanettin, Federico. 2012. *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Manchester: St. Jerome Publishing.