

**CORPUS STYLISTICS AND TRANSLATION STUDIES:
A CORPUS-ASSISTED STUDY OF JOSEPH CONRAD'S
HEART OF DARKNESS AND ITS ITALIAN TRANSLATIONS**

LORENZO MASTROPIERRO

BA, MA

**Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy**

September 2015

ABSTRACT

This thesis carries out a corpus stylistic study of Joseph Conrad's *Heart of Darkness* and four of its Italian translations. It investigates the role of textual patterns as building blocks of the fictional world and triggers of literary themes. It also investigates the effects of translation on the relation between textual patterns and the fictional world, and discusses the potential consequences of translational alterations on the text's themes.

Heart of Darkness is a complex and multifaceted text that deals with a multitude of themes and has been interpreted in many different ways. By offering an overview of the text's literary reception, I foreground two major themes that emerge from the contemporary critical debate as particularly central to the discussion about Conrad and his text: "Africa and its representation" and "race and racism". Through a keyword analysis, I establish a connection between these themes and the lexical level of the text. Adopting Mahlberg & McIntyre's (2011) model, I group keywords into categories that reflect specific aspects of the fictional world and the thematic concerns of the text. I then select groups of keywords that relate specifically to "Africa and its representation" and "race and racism" for more in-depth examination. Specifically, I analyse how the African jungle and the African natives are linguistically represented in the text. I demonstrate that repeated lexico-semantic patterns shape these fictional representations and play a fundamental part in the interpretation of the two themes related to them.

I then focus on the Italian versions and compare them in order to show the effects of translation on the lexico-semantic patterns. I show that alterations made at the linguistic level affect the interpretational level of the translations, with potential consequences for the reception of the major themes in the target context. Finally, I use computational methods to compare the original and the translations at the level of whole texts, as opposed to feature-specific comparisons. I claim that together these two perspectives provide a more nuanced understanding of the relation between source and target texts.

Through this analysis, the present thesis explores how the fictional world and literary themes are constructed and conveyed in literature and in its translation. It also

contributes to the critical discussion on *Heart of Darkness* and proposes a methodology to analyse and compare literary translations. Finally, as an interdisciplinary project, this thesis builds on the interaction between corpus stylistics and translation studies, and strengthens this relation further.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisory team. I am indebted to Michaela Mahlberg, my first supervisor, for her constant support and for pushing me to do better and better: I could not have had more inspiring guidance. I am very thankful to Maïke Oergel, my second supervisor, for the expert assistance she has offered me, and for the encouragement and optimism that meeting with her has always inspired. I would also like to thank Cecilia Goria, my third supervisor, for her advice on the Italian side of my thesis. Huge thanks must likewise go to the University of Nottingham and the School of English for their financial support, without which all of this would not have been possible.

I am thankful to my friends and colleagues who have helped me in many ways during these four years. In particular, I would like to thank Gareth, Louise, Dan, Lizzie, Jenny, Arwa, and Jacqueline for proofreading this thesis.

Finally, I am hugely grateful to Emma and my family, to an extent I would not be able to put into words here.

CONTENTS

Introduction	1
Chapter One. A corpus stylistic approach to the study of literary translation	
1. Introduction	7
2. Stylistics and the comparison of style	8
3. Corpus stylistics	11
4. Corpus stylistics at work	15
4.1 Quantitative approaches to corpus stylistics	15
4.2 Corpus stylistics and the study of textual patterns	17
5. Comparing translations	22
5.1 Form, meaning, and function in translation	22
5.2 Function and equivalence	24
5.3 Function and style	26
6. Corpora and translation	28
6.1 Corpus approaches to the study of translation	28
6.2 Corpus stylistics and translation studies	31
6.3 Computational approaches	34
6.4 Translator's style and author's style in translation	36
7. Investigating manipulation	38
8. Conclusion	43
Chapter Two. Africa and Africans in Joseph Conrad's <i>Heart of Darkness</i>	
1. Introduction	44
2. Conrad and the literary canon	45
3. The postcolonial approach in Conrad criticism	46
4. Africa and its representation	51
5. Race and racism	53
6. Conclusion	55

Chapter Three. Texts, corpora, methods: A corpus approach to literary translation

1. Introduction	57
2. The research questions	57
3. Texts and corpora	59
3.1 <i>Heart of Darkness</i>	60
3.2 The reference corpora	62
3.3 The translations	63
4. Data analysis	65
4.1 Lexis and themes	68
4.2 From words to extended units of meaning	72
4.2.1 Semantic preference and semantic prosody in <i>Heart of Darkness</i> and its translations	77
4.2.2 Lexical cohesion in <i>Heart of Darkness</i> and its translations	79
4.3 Units of meaning in translation	84
4.4 The bigger picture: Principal component analysis	87
4.5 The software	90
5. Conclusion	90

Chapter Four. From lexis to themes: The keyword analysis

1. Introduction	93
2. Generating the KW lists	94
3. The KW classification model	100
4. The KW categories	103
5. The “Africa words” and the “African words”	109
5.1 The “Africa words”	109
5.2 The “African words”	112
6. Conclusion	114

Chapter Five. “A place of darkness”: The fictional representation of the African jungle

1. Introduction	116
2. The African jungle as a “place of darkness”	117

2.1 <i>Forest, jungle, and wilderness</i>	117
2.2 The fictional representation of <i>darkness</i>	124
2.3 Africa as a metaphorical place	127
3. The fictional representation of the African jungle as a personified entity	129
4. The fictional representation of the African jungle in translation	133
4.1 The representation of the jungle as a “place of darkness” in translation	133
4.2 Lexical cohesion in the translations	138
4.3 Personification in translation	145
4.4 Shifts and manipulation	151
5. Conclusion	154
Chapter Six. “Black things”: The fictional representation of the African natives	
1. Introduction	156
2. The representation of the Africans in <i>Heart of Darkness</i>	157
2.1 Frequency comparison of the “African words”	157
2.2 The “African words” in context	159
2.3 Dehumanisation in the fictional representation of the natives	165
3. The comparison of the translations	170
3.1 The “African words” in translation	173
3.1.1 <i>Nigger/niggers</i>	173
3.1.2 <i>Savage/savages</i>	176
3.1.3 <i>Negro</i>	177
3.1.4 <i>Native/natives</i>	179
3.1.5 <i>Black/blacks</i>	179
3.2 Tendencies in the translation of the “African words”	180
3.3 Semantic preferences and prosody in translation	186
4. Conclusion	192
Chapter Seven. Applying PCA to the study of literary translation	
1. Introduction	194
2. Parallel word- and word-sequence lists	195
3. Comparing the TTs to each other	198

3.1 Comparing the TTs to each other using MFWs	198
3.2 Comparing the TTs to each other using MFSs	206
3.2.1 Comparing the frequency patterns of two-word MFSs	207
3.2.2 Comparing the frequency patterns of three-word MFSs	209
3.3 Relation between the findings	211
4. Comparing the TTs in relation to the ST	212
4.1 Section clustering in the ST	214
4.2 Section clustering in the TTs	217
4.2.1 Comparing the section clustering of the TTs to the ST	218
4.2.2 Comparing the section clustering of the TTs to each other	221
4.3 Different comparisons, different findings	221
5. Stylistic overtones in translation	222
6. Conclusion	226
Chapter Eight. Concluding discussion	
1. Introduction	229
2. Major findings	229
2.1 Textual patterns as building blocks of the fictional world	229
2.2 Textual patterns and the fictional world in translation	233
2.3 Different levels of comparison	235
3. Methodological considerations	236
4. Limitations of this study	237
5. Further research	238
References	241
Appendices	258
Appendix 1. Texts in the reference corpora	258
Appendix 2. Larger scatterplots	261
Appendix 3. Sections comparison	269

LIST OF TABLES

Table 3.1 Texts and corpora	64
Table 3.2 50 MFWs in HoD	88
Table 4.1 List 1 (HoD vs. CC; content words only)	97
Table 4.2 List 2 (HoD vs. RC; content words only)	98
Table 4.3 KWs shared by List 1 and List 2	100
Table 4.4 ‘Fictional world signal’ and ‘thematic signal’ KWs	104
Table 5.1 Comparison of the collocates of <i>forest</i>	123
Table 5.2 Comparison of the collocates of <i>jungle</i>	123
Table 5.3 Comparison of the collocates of <i>wilderness</i>	123
Table 5.4 Comparison of the collocates of <i>darkness</i>	126
Table 5.5 Collocates of <i>foresta</i> belonging to the three semantic fields	134
Table 5.6 Collocates of <i>giungla</i> belonging to the three semantic fields	134
Table 5.7 Translations of <i>wilderness</i> and <i>darkness</i> for each TT	135
Table 5.8 Collocates of translations of <i>wilderness</i> belonging to the three semantic fields	136
Table 5.9 Collocates of translations of <i>darkness</i> belonging to the three semantic fields	137
Table 6.1 Frequency of the “African words” in HoD and in the reference corpora	158
Table 6.2 Frequency of the “African words” in HoD referring to the natives	160
Table 6.3 Collocates of the “African words” belonging to the four semantic fields	163
Table 6.4 “African words” and their frequency in the TTs	173
Table 6.5 Frequency of the translations of the “African words” in each TT	180
Table 6.6 Semantic fields of the “African words” in Translation S	187
Table 6.7 Semantic fields of the “African words” in Translation B	187
Table 6.8 Semantic fields of the “African words” in Translation M	188
Table 6.9 Semantic fields of the “African words” in Translation G	188
Table 7.1 Sample of TTs parallel wordlist	196
Table 8.1 Summary of the representation of the jungle and of the natives	231

LIST OF FIGURES

Figure 3.1 Methodology overview	67
Figure 3.2 Example of unit of meaning in the BNC: <i>reduce</i>	75
Figure 3.3 The methodological circle	91
Figure 4.1 Concordance lines of <i>brooding</i>	106
Figure 4.2 Concordance lines of <i>ominous</i>	107
Figure 5.1 Concordance lines of <i>forest</i>	118
Figure 5.2 Concordance lines of <i>wilderness</i>	120
Figure 5.3 Concordance lines of <i>jungle</i>	121
Figure 5.4 Concordance lines of <i>darkness</i>	125
Figure 5.5 Concordance lines of <i>forest</i> and <i>wilderness</i> : personification	130
Figure 5.6 Concordance lines of <i>foresta</i> : personification	146
Figure 5.7 Concordance lines of the translations of <i>wilderness</i> : personification	148
Figure 6.1 “Physicality” semantic field	161
Figure 6.2 “Collectives” semantic field	161
Figure 6.3 “Incomprehensibility” semantic field	162
Figure 6.4 “General negative” semantic field	162
Figure 6.5 Concordance lines of <i>black</i> used as an adjective	164
Figure 6.6 Collocates of <i>negro</i> (left side) and <i>nero</i> (right side)	175
Figure 6.7 Collocates of <i>persona di colore</i>	179
Figure 7.1 Score plot based on the 250 MFWs of the TTs	201
Figure 7.2 Scatterplot based on the 250 MFWs of the TTs	202
Figure 7.3 Score plot and scatterplot based on the 350 two-word MFSs of the TTs	208
Figure 7.4 Score plot and scatterplot based on the 350 three-word MFSs of the TTs	210
Figure 7.5 HoD sections score plot based on 250 MFWs	215
Figure 7.6 HoD sections scatterplot based on 250 MFWs	216
Figure 7.7 TT sections score plots based on 250 MFWs	218

INTRODUCTION

This thesis presents a corpus-assisted analysis of Joseph Conrad's ([1902] 1994) *Heart of Darkness* (HoD) and four of its Italian translations, combining corpus stylistics and descriptive translation studies. Corpus stylistics is the application of corpus methods to the study of literature by relating linguistic description with literary interpretation (Mahlberg 2013). Descriptive translation studies is concerned with the empirical examination of the product, the process, and the function of translation, and its theoretical and applied implications (Toury 1995). Sharing a number of fundamental assumptions, these two disciplines are particularly apt to be combined. First, they both focus on language in use and consider language as a social phenomenon, closely connected with its context of production. Second, this context of production takes centre stage in both disciplines. Translating, as well as writing and reading literature, is a highly contextualised activity, related to linguistic and extra-linguistic aspects alike. Both corpus stylistics and descriptive translation studies seek to account for the extra-linguistic aspects in the examination of the linguistic ones. Finally, the two disciplines hinge on comparison. Corpus stylistics defines and describes style on the basis of the identification of a difference between a sample and a norm, a text and a reference corpus (Halliday 1971, Fowler 1966, Leech & Short 2007). Similarly, descriptive translation studies rests on the comparison of the source text (ST) and the target text (TT) as the primary method of analysis (Bassnett 2007, Munday 2008). Multidisciplinarity is therefore a key feature of this thesis.

There are many advantages in adopting a corpus approach in the study of literature. Computer technologies allow the researcher to analyse texts on a scale and level of statistical detail that would be unachievable for the human analyst alone. Novels can be as short as HoD, which is about a hundred pages, and still it would be impossible to parallel the computational abilities of the computer with standard reading. The availability of data also makes the stylistic analysis more empirical, as linguistic evidence can be used to substantiate claims about style (Leech & Short 2007: 38). However, it is not only a matter of sheer computing. Corpus methods offer new and unprecedented ways to study the text. They add new descriptive tools to the

stylistician's toolkit and make possible types of analysis that were not previously possible. With new methods come new research questions.

These advantages equally apply to the corpus linguistic study of translation. Corpus methods provide quantitative data about the ST and the TT, to an extent that is not otherwise obtainable. This data can be used to support textual comparison empirically and to improve its accuracy. Comparison itself is thus enhanced by the use of the computer. Finally, new analytical possibilities generate new research questions, questions that can be asked and answered only by the use of corpora. In this thesis, I will use a corpus approach to study literature and its translation in conjunction, showing how these advantages make this approach the most suitable to achieve the aims of the present study.

This thesis has three over-arching aims. The first aim is to carry out a corpus stylistic study of HoD, focusing on the role of textual patterns as building blocks of the fictional world. I argue that literary themes are related to the way the fictional world is linguistically constructed; therefore, examining the linguistic construction of this fictional world will allow me to study how textual patterns contribute to the text's thematic concerns. The analysis will concentrate mainly on lexical and semantic patterns, through the examination of semantic preferences and semantic prosodies. Semantic preferences and prosodies play a key part in conveying meaning in texts (Sinclair 1991, 2004), and their role in literature has been proved to be similarly meaningful (Louw 1993, Fischer-Starke 2010, Munday 2011). Moreover, semantic preferences and prosodies are not defined on the basis of frequency cut-off points, i.e. their identification does not depend upon the frequency of the items that form them. This is important because HoD is a short text, especially compared to the amount of words that corpus linguistics usually deals with. Semantic preferences and prosodies are not affected by the size of HoD, as they account for the cumulative effect of words that do not occur frequently in the text: there is no frequency threshold for a word to contribute to a semantic field. In contrast, the identification of frequency-based patterns (such as collocational or colligational patterns) can be affected by the size of the text/corpus, given that these patterns are usually defined in statistical terms. As a result, words that are thematically significant but do not occur very often are unlikely to show any relevant collocational or colligational patterns, whereas semantic preferences and prosody can still be identified in relation to low-frequency items.

The second aim of this thesis is to compare four Italian versions of HoD, concentrating on the effect of translation on the relation between textual patterns and the fictional world. I will show that this relation is affected by the linguistic alterations resulting from translating, independent of whether the alterations are conscious translation strategies or unconscious changes. Modifications in the patterns of the original result in discrepancies in the representation of the fictional world between the ST and the TTs. This study will argue that these discrepancies are not only limited to the linguistic level, but are also concerned with the interpretational level. Specific aspects of HoD's fictional world are strictly related to its major themes; therefore, the changes in the patterns can also affect the interpretation of these themes.

Finally, the third aim of this thesis is to establish a methodological synergy between corpus stylistics and translation studies, outlining a framework with broader applicability beyond the specific analysis of Conrad and its Italian translations. Nevertheless, the present study is mainly focused on Conrad and HoD, as this text and its translations represent an ideal case study to explore the relationships between textual patterns, fictional worlds, and literary themes.

There are many reasons behind the decision to focus on Joseph Conrad and HoD. First of all, Conrad is an author that, from a stylistic point of view, has a lot to offer. With his writing, Conrad aims “to make [the reader] hear, to make [him/her] feel – it is, before all, to make [him/her] *see*” (Conrad 1950: ix, emphasis in the original). As he puts it in his literary manifesto, the renowned preface to *The Nigger of the Narcissus* (Conrad [1897] 1950), Conrad believes that a writer should be “devot[ed] to the perfect blending of form and substance” (Conrad 1950: ix), should put a “never-discouraged care for the shape and ring of sentences” (Conrad 1950: ix). He is therefore very aware of the importance of the linguistic form in his literary aesthetic, knowing that it can assist in the development of his artistic aims. Conrad is well-known for going through considerable effort to pursuit *le mot just* (Knowles 2009b: 34). This attention to prose form as a means through which to convey his artistic intentions makes Conrad an exemplary subject for a study such as the present one, which seeks to examine how linguistic features contribute to the fictional world and to major literary themes.

Secondly, Conrad and especially HoD occupy a central position in the English literary canon, and their popularity is reflected in over a century of sustained interest. This is very useful for a corpus stylistic analysis because it means that there is an abundance of evidence of how both readers and critics have responded to Conrad's novels. The study can therefore be informed by, and interact with, this critical background. This also applies to the translations. Conrad's popularity is by no means tied to national boundaries, but is instead widespread all over the world. The critical discussion of his works has been equally lively in Italy (Ciompi 2005), and HoD in particular is the novel with the largest number of Italian translations in Conrad's oeuvre (Curreli 2009: 149), with over 20 versions since the first one in 1924. As such, HoD is a valuable text for the study of literary translation. An additional benefit to studying Conrad is that the vast majority of his texts are readily available on the internet, free from copyright restrictions. This means that the process of retrieving texts in the appropriate digital format and building reference corpora is greatly facilitated.

This thesis is structured as follows. Chapter 1 sets the theoretical ground for the interdisciplinary dialogue between corpus stylistics and translation studies. It introduces the study of style with corpus methods, both in literature and in translation, and illustrates the specific approach that this work adopts. Chapter 2 defines the two major themes of HoD that this thesis focuses on: "Africa and its representation" and "race and racism". To do so, the chapter outlines the critical reception of HoD and emphasises the themes that seem to be the most relevant to contemporary criticism. Chapter 3 sets out the methodology of this thesis and discusses the practical implication of using a corpus stylistic approach to study literary translations. This chapter also details my research questions, explaining how each one is addressed by my methodology.

In Chapter 4, I start my analysis, using keywords to establish a link between the lexical level of HoD and the two major themes identified in Chapter 2. By grouping the keywords into 'fictional world signals' and 'thematic signals', this chapter identifies categories of relevant items that will be analysed in detail in the following chapters. Chapter 5 focuses on the first group of keywords, related to "Africa and its representation". In this chapter, I carry out a quantitative and qualitative analysis of the fictional representation of the African jungle, both in the ST

and in the TTs. I discuss the role that lexico-semantic patterns play in building up the representation of the jungle, and examine the effect of translation on how the jungle is depicted in the TTs. Chapter 6 concentrates on the second group of keywords, which is linked to the theme of “race and racism”. This chapter uses a similar approach to the previous one, this time applied to the study of the fictional representation of the African natives. The way the African characters are linguistically constructed in both the ST and the TTs is examined and compared, and the discrepancies identified are analysed in terms of the effect they have on how the theme of race and racism is interpreted.

Chapter 7 approaches the texts from a different perspective to the previous chapters, as it aims to compare the translations as whole texts, rather than focusing on specific linguistic features within the texts. I do this by employing principal component analysis, a statistical method that offers a measure of the general degree of similarity and difference between the texts, allowing me to relate the results of the specific comparisons in Chapters 5 and 6 to the overall picture of the text relations provided by this chapter. Chapter 8 discusses the central findings of this thesis, highlighting my contributions to the field, as well as outlining some further avenues for future research.

Throughout this thesis, I use the following typographical conventions. Italics are used for language examples, both individual words and short sentences. Small capitals are used for lemmas. Single quotes (‘’) are used to introduce terminology, while double quotes (“”) are employed for direct quotations, categories/groups, “so-to-speak” uses, and translations. The translations provided in inverted commas are always literal, as opposed to “literary”, translations.

Overall, this thesis aims to make an innovative contribution to the application of corpus stylistics to the study of literary translation. By examining the relation between textual patterns, the fictional world, and central literary themes, I contribute to a more nuanced understanding of the interaction between form and meaning in literature, and how translation affects the interplay of the two. In doing so, I seek to demonstrate the potential of interdisciplinary research, by triangulating corpus methods, stylistics, and translation studies. This thesis also aims to make a contribution to the critical discussion of Conrad and HoD. My focus on central themes allows me to contribute to key areas of Conradian studies, offering a linguistic

perspective on fundamental aspects of HoD. Fredrick R. Karl (in Pursell 2009: 83) has claimed that Conrad is a writer “about whom very little remains to be said”. Ultimately, I seek to challenge this narrow viewpoint by demonstrating how an interdisciplinary approach to Conrad’s HoD foregrounds previously overlooked textual features that in turn provide fresh insight into such a popular and iconic text.

CHAPTER ONE

A corpus stylistic approach to the study of literary translation

1. Introduction

This thesis discusses the relationship between lexico-semantic patterns and fictional worlds, both in literary texts and in their translations. The approach used is that of stylistics, while corpus linguistics provides the tools for the analysis. This chapter will introduce the interaction between the two, presenting the broader fields in which my research places itself. Stylistics will be introduced in Section 2. In particular, I will emphasise its focus on both linguistic description and literary appreciation and its comparative nature. In Section 3, I will discuss the interface between stylistics and corpus linguistics. I will explain why corpus stylistics is the most suitable approach to study HoD and its Italian translations. At the same time, I will also consider more broadly the application of corpus methods to the study of literary texts in general, reviewing the overall development of the field. In Section 4, I will discuss the existing research in corpus stylistics, mainly focusing on those approaches that share features with this thesis. In Sections 5 and 6, I will introduce translation into the picture. Analysing style in translation often involves the comparison of the ST and the TT; I will argue that corpus stylistics can enhance this comparative perspective, helping to answer questions in translation studies with the support of empirical data. Moreover, Sections 5 and 6 will stress that a (corpus) stylistic approach to the study of translation is not necessarily tied to linguistic description, but is instead able to account for extra-textual factors as well. In this respect, the present analysis will look at potential manipulative effects created by alterations to the ST's linguistic patterns in the TTs. The notion of manipulation and its investigation in translation studies will be presented in Section 7. Finally, in Section 8 I will close the chapter with some concluding remarks.

According to Simpson (2004: 8), stylistic analysis has “no necessarily ‘natural’ starting point”, but rather it is flexible and circumstantial, in accordance with the aspects of language we want to concentrate on. There is no aspect of language whose analysis has priority over other aspects; rather, different levels interact,

complement, or even collide with each other. Therefore, depending on the aspects of language taken into account in the analysis, one may want to discuss a particular method or concept instead of another. This flexibility is reflected in the structure of this and the following chapters. In line with Simpson's (2004) conception of stylistics, this chapter aims to provide the broader setting of the present analysis, leaving to the following chapters the discussion of the specific concepts used in each of them. The discussion here deals mainly with the key points that ground the interaction between stylistics, corpus linguistics, and translation. In other words, this chapter is concerned with the "big picture", while the various "starting points" will be discussed later on.

2. Stylistics and the comparison of style

In his "Closing statement: Linguistics and poetics", Jakobson (1960: 377) argued for "the right and duty of linguistics to direct the investigation of verbal art in all its compass and extent", claiming that there is no reason to "separate the literary from the overall linguistic". Today, Jakobson's (1960) claims are widely accepted and the separation between literary and non-literary language does not stand up (Carter 2004). Linguistic models and techniques are in fact extensively applied to analyse literature. Stylistics, the linguistic study of style (Jeffries & McIntyre 2010, but also Leech & Short 2007, Lambrou & Stockwell 2007), aims to explain the relation between language and literary effects. What role does the verbal form of a text play in our perception of its literariness? How does the author use his/her language to achieve creative effects? These are some of the questions stylistics tries to answer. However, the goal here is not just describing language usage in literature. Rather, stylistics uses linguistic description to enhance our understanding of the text, to help its interpretation.

In addition to examining linguistic features of literary texts, stylistics is also concerned with studying the part these features play in our interpretative experience. Linguistic description and literary appreciation are equally important. This idea of stylistics has been advocated by Spitzer (1962), who proposes the 'philological circle' to illustrate the interaction between linguistic description and literary appreciation. The two are linked to each other in a cyclical motion; there is no priority given on

either but rather mutual collaboration. One motivates and informs the other. The text is considered at the same time a sample of language in use and a literary work, and is analysed as such.

Stylistics is therefore a multidisciplinary field by definition, encompassing perspectives from literary studies and linguistics alike. However, it has been criticised by both linguists and literary critics. Stubbs (2005: 5), for example, explains that linguists are often sceptical towards stylistics, as they are uninterested in investigating individual texts. On the contrary, they tend to study regularities across language, in order to develop general theories. On the other hand, literary scholars can find stylistic analysis flawed by circularity and arbitrariness (Fish 1979, 1980), especially when claims on literary style are based on a few specifically selected linguistic data. But despite the scepticism, the major strength of stylistics is exactly the interaction of two separate yet strongly linked disciplines. Far from being the only focus of the analysis, the linguistic description complements the critical reading. It helps to explain why we perceive the text in a given way and offers an observable justification for our readerly experience, which might otherwise remain unexplained. Even if this served just to confirm existing interpretations with empirical data, it would be a positive outcome nevertheless. In fact, as Stubbs (2005: 6) explains, this would give us confidence that the stylistic method is reliable, because it matches the findings of other broadly established approaches. Likewise, when literary appreciation informs linguistic scrutiny, the analysis acknowledges the ‘literariness’ (Carter 2004) of the text, so that each formal feature is studied in the context of the artistic effects conveyed by the text.

Besides its multidisciplinaryity, another characteristic feature of stylistics is that it is “essentially comparative in nature” (Halliday 1971: 341). Investigating the style of a text often entails identifying features that characterise that text compared to others. For instance, Senn (1980) recognises in Conrad’s narrative an adjectival style. This implies that Conrad’s use of adjectives is perceived as different from what we would expect. If every other author used adjectives in the way Conrad does, than this feature would not be characteristic anymore. Therefore, in order to classify Conrad’s use of adjectives as a stylistic feature of his novels, we need to compare them with other novels by other authors and recognise a difference in their adjectival use. In

other words, “usage in one text is significant only by comparison with usage outside it” (Fowler 1966: 22).

Comparison can be established at a number of different levels. Leech & Short (2007) identify the following three principal ones. First, a textual example can be compared to a general set of texts, which represent a heterogeneous sample of language (e.g. comparing Conrad’s novels with other general purpose texts). Second, a textual example can be compared to a set of texts by the same author or in the same genre (e.g. comparing HoD with the rest of Conrad’s novels). Third, a textual example can be compared to the whole text from which it is taken (e.g. comparing the first chapter of HoD with the remaining ones). These different comparisons relate to three types of deviation from a perceived norm: primary deviation, secondary deviation, and internal deviation (Leech & Short 2007: 44). ‘Primary deviation’ is a deviation from general language; ‘secondary deviation’ is a deviation from authorial or genre-specific norms; ‘internal deviation’ is a deviation from the norms of the text itself (Leech 1985: 45-48). In the present study, I will focus mainly on secondary deviations, comparing HoD against Conrad’s other novels and against other literary texts from the same period.

I will also compare the translations of HoD, one against the other and against the ST. In translation studies, as in standard stylistic analysis, comparison plays a fundamental role, considering that most of the questions the discipline deals with – from the strategies employed by the translator to the different status of the ST and TT in their respective cultures – are analysable mainly through comparing translation against original, or vice versa. “More broadly”, Bassnett (2007: 19) explains, the comparison of the original and the translated text “will expose the relationship between the two cultural systems in which those texts are embedded”. In terms of stylistic analysis, it is only through the comparison of the ST with the TT that differences in style can be discovered: “Any alteration, muffling, exaggeration, blurring, or other distortion of authorial voice will remain hidden until and unless some element of the TT reveals the mediation or until the TT is compared to its ST” (Munday 2008: 14).

For the present research, then, comparing texts is the most important analytical procedure. It is not only an intrinsic part of the stylistic analysis, but it is also fundamental in studying translation. Given these premises, in the next section I will

explain why a corpus approach to the study of style is the most appropriate when comparison takes centre stage.

3. Corpus stylistics

In their second edition of *Style in Fiction*, Leech & Short (2007: 286) acknowledge that stylistics has experienced a “corpus turn” in recent years, with a growth in the number of studies applying corpus methods to the analysis of literature. The increasing availability of texts in electronic format and user-friendly software for text analysis has made it possible for more and more researchers to benefit from the advantages that this application entails. In particular, there are two main areas in which these advantages are most striking: providing quantitative data and new analytical techniques.

In the previous section, I have argued that stylistics is essentially comparative in nature (Halliday 1971). Comparing is necessary to prove that a linguistic feature characterises one text against the background of standard usage. There must be a difference in the consistency and tendency of occurrence compared to a norm in order to define that feature as typical of a given author or text. Consistency and tendency are best examined in terms of frequency (Leech & Short 2007: 34), so it becomes clear that quantitative data are important to the analysis of style. All instances of recognisable style, whether of an individual writer, of a single book, or of an entire genre, have distinctive features that can be analysed in terms of relative frequency (Halliday 1971: 343). Through frequency comparison it is possible to establish whether a feature is distinctive – i.e. ‘deviant’ – or not, namely whether a difference exists “between the normal frequency of a feature, and its frequency in the text or corpus” (Leech & Short 2007: 39). Quantitative description is therefore essential to any claim about style: “The more I, as a critic, wish to substantiate what I say about style, the more I will need to point to the linguistic evidence of texts; and linguistic evidence, to be firm, must be couched in terms of numerical frequency” (Leech & Short 2007: 38). A corpus approach to the study of style enhances the quantitative description of the text and provides numerical data to a level of statistical precision that would be impossible for the researcher to achieve manually. Computers retrieve

quantitative figures faster and more accurately than any human reading. This makes it possible to analyse exhaustively large amount of texts, such as very long novels or even the whole oeuvre of an author, in a way that would not be feasible to do manually.

However, it would be a simplification to argue that corpus methods reduce the notion of style to quantification alone. There is no such thing as the completely objective measurement of style. In fact, not all deviant features have stylistic significance, as numbers do not tell us whether a linguistic aspect has “value in the game” (Halliday 1971: 344) or not, that is, whether or not it is ‘foregrounding’. A foregrounded feature is a feature that “relates to the meaning of the text as a whole” (Halliday 1971: 339) and is used for a specific literary end (Leech & Short 2007: 40). A linguistic feature can occur unusually frequently and yet not contribute in any way to the style of a text. Statistics alone do not prove style. Yet, although not all deviant features are foregrounding, all foregrounding features are instances of statistical deviance (Leech & Short 2007: 41). Therefore, as far as quantification is possible, the adoption of corpus methods for the study of style is extremely beneficial.

The second major advantage of applying corpus methods to stylistic studies is the introduction of new analytical techniques. The use of the computer not only provides assistance for the stylistic analysis, but enriches it with new practices and descriptive categories. Besides making the search for stylistic features faster, more precise, and more reliable, corpus methods expand altogether the range of features identifiable. Mahlberg (2013: 18) explains that the innovations that have come about in corpus linguistics have in turn extended the stylistician’s toolkit. Borrowing models and concepts from the cognate field of corpus linguistics, the stylistician can apply descriptive categories to the study of style that are typical of a corpus approach and cannot be studied otherwise.

Consider ‘keywords’ and ‘semantic prosody’, for instance, which are two important concepts employed in this thesis. Keyword analysis does not simply provide evidence for what one might think a text is about, but also reveals features which might be difficult to notice without the help of the computer. For example, Culpeper (2009) uses keyword analysis on the character-talk from Shakespeare’s *Romeo and Juliet* and shows that this corpus technique not only confirms what we already expected about the characters of the play (“that Romeo is all about love”,

Culpeper 2009: 53). It also points out grammatical features of Juliet's talk (*if, yet, or, would, be*) that, at first, might not seem relevant, but that actually distinguish stylistically Juliet as a character experiencing anxieties (Culpeper 2009: 53).

Louw (1993) shows that semantic prosodies can play an important role in establishing irony and claims that the analysis of semantic prosody should be at the heart of what he calls 'radical stylistics' (Louw 1993: 173). Departing from or adhering to an established semantic prosody has an effect on the reader, who will have his/her prior knowledge about the use of an item challenged or not, respectively. Louw (1993: 161) shows how the poet Philip Larkin makes use of the bad semantic prosody of *utterly* to convey a negative meaning in the final line of his poem "First Sight". On the contrary, David Lodge, in his novel *Small World*, departs from the established bad semantic prosody of the item *bent on* to create an ironic effect. Despite its role in imbuing an item with "a consistent aura of meaning" (Louw 1993: 157), we are mostly unaware of semantic prosodies, because they escape human intuition (Louw 1993: 157); yet they can be uncovered effectively through the analysis of corpora. The study of keywords and semantic prosodies – discussed in detail in Chapter 3 – has been made possible through the adoption of a corpus approach, offering a perspective on the texts which traditional stylistics alone could hardly obtain.

The advantages so far described make 'corpus stylistics', i.e. the corpus-assisted study of literary style, a more suitable approach for the present research than "traditional" stylistics. First, in order to study the relation between themes and linguistic features, HoD and its Italian translations need to be examined throughout, as major themes are likely to be spread all over the novel. A corpus stylistic approach allows me to search the whole texts in a faster and more accurate way. Even though HoD is a short novel, a similar analysis would not have been possible without the aid of a computer. Second, this research is comparative in nature, like most research in stylistics and translation studies. A corpus stylistic approach enhances the comparison of texts and provides me with numerical data with which to substantiate my claims about style and differences of style in the translations. Finally, corpus stylistics widens the range of phenomena I can look at and facilitates the identification of consistent and repeated patterns in texts. Some of these phenomena are notoriously invisible to the naked eye: for example, semantic prosodies can develop from the use

of individual words, whose cumulative effect is noticeable only if they are taken into account all together. I will show that, although “inaccessible to human intuition” (Louw 1993: 157), semantic prosodies play a major role in the process of building up the fictional world.

So far, the discussion has emphasised the technical advantages that a corpus approach brings to stylistics. This might have given the impression that corpus stylistics is mainly interested in quantification and description, as opposed to more qualitative analysis. This is not the case. On the contrary, in corpus stylistics the qualitative perspective is at least as important as the quantitative one. Quantification alone cannot explain a style; rather, it can support the process of interpretation. Whereas “traditional” stylistics relies on textual evidence and linguistic tools to substantiate subjective interpretations of literary texts, corpus stylistics does the same, but with more textual evidence and more (corpus) linguistics tools. In other words, quantification and description are never the goal of the study, but rather a means to formulate or test our intuition-based interpretations (Ho 2011: 10).

This perspective is reflected in Mahlberg’s (2013: 5) successful definition of corpus stylistics, that it is “the application of corpus methods to the analysis of literary texts by relating linguistic description with literary appreciation”. Mahlberg (2013) draws on Spitzer’s (1962) ‘philological circle’ to explain the role of the corpus approach in the relation between linguistic description and literary appreciation. Corpus methods do not simply substitute the linguistic description. They are in fact suitable to describe some linguistic phenomena, but not all. However, the linguistic description alone can only observe certain features, but not others. Corpus linguistics has therefore provided new tools to assist the linguistic description and explain a wider range of phenomena. Finally, the two complement each other and together inform and are informed by literary appreciation. Thus, the ‘corpus stylistic circle’ has three independent but mutually related components, instead of two: literary appreciation, linguistic description, and corpus linguistic description (Mahlberg 2013: 12). This emphasises the importance of the corpus method in stylistic analysis, showing that its addition is more than a mere accessory.

4. Corpus stylistics at work

Although a sub-discipline in itself, describing corpus stylistics as a unified field would be misleading. On the contrary, it is exceptionally varied, and this variety is reflected in a wide range of different research, more or less quantitative depending on the method and approach used. This section is therefore not intended to offer a comprehensive view of the field as a whole, but rather will focus on research that touches upon the main concepts the present work is concerned with. In particular, in Section 4.1, I will concentrate on more quantitative approaches to corpus stylistics, such as computational stylistics and stylometry. In Section 4.2, I will discuss more qualitative approaches to corpus stylistics, focusing specifically on the use of corpus methods to study patterns in texts and in literature.

4.1 Quantitative approaches to corpus stylistics

To start with, it is important to state that the use of the computer to assist analyses of literary texts predates the term ‘corpus stylistics’ itself. Computer-assisted methods have been extensively used in the fields of computational stylistics and stylometry long before corpus stylistics was established as a field in its own right. For example, Burrows (1987) uses computational techniques to study Jane Austen’s language and the idiolects of her characters. More recent but similar in scope is Craig’s (2008) examination of the different styles in the dialogues of fifty of William Shakespeare’s characters, by looking at the frequency patterns of the fifty most frequent words in these dialogues.

Computational methods have not only been used to analyse the writing of specific authors, but also to explore style variation across fiction. Egbert (2012) builds on Biber’s (1988) original multi-dimensional analysis and applies it to a large corpus of British and American nineteenth century fiction. Whereas Biber (1988) examines the distribution of 67 linguistic features across different registers to analyse language variation in speech and writing, Egbert (2012) focuses on literary style and stylistic variation across 100 texts of 10 different writers. He takes into consideration the degree of co-occurrence of 59 linguistic variables to determine stylistic tendencies

between “thought presentation vs. description” (Dimension 1), “abstract exposition vs. concrete action” (Dimension 2), and “dialogue vs. narrative” (Dimension 3). The resulting variation in these dimensions across the texts proves to be accurate in distinguishing the writing style of the novels, both at a between-author and at a within-author level.

Authorship attribution is another area of study in which computational methods are frequently applied to literary texts. For instance, Binongo & Smith (1999) explain the use of principal component analysis in authorship attribution through the examination of Shakespeare’s *Pericles*, the first two acts of which were argued to be written by George Wilkins rather than by Shakespeare himself. Rybicki et al. (2014) apply ‘Rolling Delta’, an implementation of Burrow’s (2002a) Delta authorship attribution algorithm, to the collaborative works by Conrad and Ford, i.e. *The Inheritors*, *Romance*, and *The Nature of a Crime*. Rybicki et al. (2014) are able to show which parts of each text display a higher or lower agreement with the style of either Conrad or Ford (represented by their respective, single-authored, works). In this way, they signal the authorship of the passages. Hoover (2002) carries out a cluster analysis to investigate whether word sequences are more effective than individual words in disambiguating texts by different authors and grouping together texts by the same writer. Using a corpus of the first 30,000 words of 29 novels by 17 authors, he proves that a cluster analysis with word sequences as variables attributes authorship more accurately than using single words.

These studies are characterised by a strong emphasis on the quantitative side of the analysis, making use of advanced computational and statistical techniques, while the present study aims to balance quantitative and qualitative investigation. This does not mean that computation has nothing to offer to this thesis. In Chapter 7, principal component analysis will be used to compare HoD and its translations as whole texts, rather than on the basis of individual features (see Chapter 3 for more details). Nevertheless, this thesis as a whole does not follow the computational stylistic approach to quantitative data. Rather, my research is set in a different branch of computer-aided studies of literature, where qualitative analysis plays a much more significant role. As suggested by Mahlberg (2013: 178), the use of technically less complex computational and corpus methods in corpus stylistics puts special emphasis on the qualitative aspects of the analysis. I have mentioned before that ‘corpus

stylistics' is an umbrella term that covers a wide range of methods and approaches, but it will be used here to distinguish it from the computational approach discussed so far. In the next section, I will discuss relevant corpus stylistic research.

4.2 Corpus stylistics and the study of textual patterns

The present study looks at lexico-semantic patterns in HoD and its Italian translations because patterns are a constituent element of all language registers and play a fundamental role in the conveyance of meaning. 'Patterns' in language can be defined as any repetition of words, sounds, rhythms, or structures (Hunston 2010: 152). It is the repetition itself that gives meaning to the patterns, which goes beyond that of their individual elements. Sinclair (1991, 2004), with his 'unit of meaning' model (discussed in detail in Chapter 3), has shown that meaning is conveyed also by patterns, rather than being associated with single elements only: meaning affects structure and structure affects meaning (Sinclair 1991: 496). Of course, this is equally true for the language of literature, where patterns have been seen to develop many different functions.

In Section 3, I have suggested that corpus stylistics helps to identify repeated patterns that otherwise would remain unnoticed. Stubbs (2005: 15), for example, argues that looking at patterns, rather than single words, is more effective if one wants to understand the complaints about repetitiveness that many critics have directed at HoD. To prove so, he reports Youmans's (1990) figures of lexical variation across different writers. HoD's proportion of different words over the total number of running words is not very different from the proportion displayed in *Middlemarch* by George Eliot or *Oliver Twist* by Charles Dickens, for instance. In contrast, Stubbs (2005: 15-16) explains that repetitiveness in HoD arises from the repeated use of lexico-grammatical patterns, such as long strings of adjectives and nouns, "abstract noun + adjective with a negative prefix" constructions, or more general words with negative prefixes and suffixes. These patterns would not have emerged if only single words had been taken into account.

Similarly, Starcke (2006) focuses on patterns in her analysis of Jane Austen's *Persuasion*. She looks at phraseology because multi-word units are more frequently

units of meaning than single words (Starcke 2006: 88) and shows that two of the most recurrent 3-grams (*she could not* and *she had been*) develop semantic and grammatical patterns that can shed new light on the analysis of the characters related to them. Starcke's (2006: 102) final aim is to demonstrate that the corpus analysis of literary texts can offer a valuable contribution to the understanding of a novel even after more than two hundred years of literary criticism. The range and scope of Starcke's (2006) analysis are extended in Fischer-Starcke (2010), which looks at semantic and syntactic patterns contributing to the encoding of literary meanings in Jane Austen's *Northanger Abbey*. She argues that not only individual items link to each other in patterns to encode meaning, but also more complex patterns and structures are interrelated to create larger networks (Fischer-Starcke 2010: 196). In this respect, corpus stylistics makes it possible to analyse longer stretches of text in linguistic detail – revealing the links existing between different parts – more than what a literary critical analysis can do (Fischer-Starcke 2010: 197). Ho (2011) proves that the aid of corpus methods in retrieving and analysing patterns is effective also in areas of language which are notoriously difficult to treat quantitatively, such as figurative language. She questions Van Peer's argument that “no computer program can automatically disambiguate figurative language as such” (in Ho 2011: 202) explaining that corpus techniques are not meant to “automatically disambiguate” figurative language. Rather, the tool makes it easier, or possible altogether, to see patterns in figurative language (Ho 2011: 202).

It is important to note that the identification of patterns with a corpus approach is not only a matter of being able to search across large amounts of language at once, highlighting features not visible through the analysis of limited text extracts. A major advantage comes about with the sorting capability of the software, which displays language in a way that is convenient for the detection of repeated patterns, e.g. concordance lines (see Chapter 3 for more details). Concordance lines allow the corpus to be read vertically (Tognini-Bonelli 2001: 3), scanning for repeated patterns across several texts at the same time in a way that would not be possible reading a text “horizontally”. Concordances are therefore a very useful tool for observing patterns in language. The quality of evidence they offer is “superior to any other method” (Sinclair 1991: 42), for they gather objectively all the occurrences of any given item and sort them alphabetically to make it much easier for the researcher to spot patterns.

As such, concordances are recurrently used in corpus stylistics. An example is Mahlberg & Smith's (2010) study of Jane Austen's *Pride and Prejudice*, in which they focus on the concordance lines of *civility* and *eyes*. The fact that concordance lines foreground strikingly verbatim repetition of the same words makes it possible to effectively identify repeated lexical patterns that convey meaning. This is the case of LIFT *her eyes*, a structure that is used to reveal the character's inner feelings without stating them directly: *lifted up her eyes in amazement*, *unable to lift up her eyes*, or *dared not to lift up her eyes* are all expressions of the characters' emotional state conveyed in terms of body language.

One of the functions that patterns can develop in literature which is particularly significant for the present research is that of acting as building blocks of fictional worlds. Mahlberg (2013: 3) illustrates how 'clusters', repeated sequences of words, can externalise features of both the fictional world and of the characters in it. To do so, she retrieves and classifies 5-word clusters on the basis of formal features in a corpus of Dickens's novels. The resulting cluster categories ("labels", "speech clusters", "body part clusters", "*as if* clusters", and "time and place clusters") are then related to specific functions. For example, "label" clusters are shown to contribute to the identification of characters and themes, "*as if* clusters" signal the author's or narrator's comments and interpretation, and "time and place clusters" define the "when" and "where" of the fictional world (Mahlberg 2013: 40).

The importance of patterns as building blocks of characterisation is also discussed in Mahlberg & Smith (2012), Mahlberg (2012), and Hori (2004). Using the web tool CLiC (Corpus Linguistics in Cheshire¹), Mahlberg & Smith (2012) analyse patterns in suspended quotations and argue that they contribute significantly to shaping characters in Dickens's *Hard Times*. Particular attention is put on the patterns of *pause* (*after a pause*, *after a short pause*, *after a moment's pause*), which are shown to be overused in suspensions compared to their use outside suspensions. These *pause* patterns add up to the description of the characters' body language and, at the same time, create an actual pause in their speech, in this way enacting a twofold function in which style and narration are related. Mahlberg (2012) and Hori (2004) focus on demonstrating how given patterns can be related to specific characters. Mahlberg (2012) looks at body language clusters linked to Tulkinghorn and Bucket in

¹ CLiC is freely available at <http://clic.bham.ac.uk/> (last accessed December 2015).

Dickens's *Bleak House*. Hori (2004) looks instead at collocational patterns in Dickens's language and demonstrates how different recurrent collocations distinguish characters, not only in their speech, but also in non-dialogues used to describe their specific appearance, behaviour, or actions.

The role of formal patterning in the process of characterisation is in line with a cognitive stylistic understanding of the phenomenon. According to Culpeper's (2002) model, the impression of characters in the reader's mind is the result of the interaction of two elements: the reader's prior knowledge of people in the real world and actual textual cues in the pages of the book. The reader picks "characterisation cues from linguistic triggers" (McIntyre 2014: 152) and processes them together with their prior world knowledge in order to produce their understanding of a given character. Mahlberg (2013, 2012) has shown that linguistic patterns occurring consistently throughout a text can be part of the bottom-up textual information, that is, the linguistic triggers that, together with prior knowledge, shape the impression of a character in the reader's mind.

In addition to their function as building blocks of fictional worlds, and as a result of that, patterns can also contribute to major themes in literary texts. The present thesis intends to examine this contribution, analysing lexico-semantic patterns in HoD and in its Italian translations. The role of patterns to the process of theme construction in HoD has been briefly discussed by Stubbs (2005). He argues that repeated phrasal patterning and recurrent lexical fields not only contribute to the feeling of repetitiveness in Conrad's short novel, but also convey the theme of uncertainty that characterises both Marlow's narration and the whole text. Stubbs (2005: 19) refers specifically to the "mist/haze" semantic field, to the *as if/as though* constructions, and to the "preposition + *the* + noun + *of* + determiner + noun" phrases (*in the midst of the incomprehensible, into the gloomy circle of some inferno, into the heart of the darkness, etc.*), explaining that their frequent occurrences blur the boundaries between appearance and reality, certainty and uncertainty. However, Stubbs (2005) does not discuss in detail the precise function of these patterns and hints only at the role they may play in the interpretation of the short novel. In my analysis, I will instead show in depth how specific lexico-semantic patterns that build up fictional entities also participate in the process of theme construction. I will also discuss the part these patterns can have in shaping the readerly experience, paying

particular attention to the relation between linguistic features and established literary readings.

It is important to note that the identification of patterns in itself does not constitute a stylistic analysis. Patterns do not necessarily represent a stylistic feature, as much as frequency does not always indicate a style. It is ‘foregrounding’ that makes a feature stylistically relevant, and a foregrounding feature – a feature that “relates to the meaning of the text as a whole” (Halliday 1971: 339) and is used for a specific literary end (Leech & Short 2007: 40) – can be recognised as such mainly through qualitative analysis. This stresses once more the importance of balancing the quantitative and qualitative perspective in a corpus stylistic study, so that the linguistic description of patterns is related to the interpretation of the text. Mahlberg (2012: 93) maintains that in the corpus-assisted analysis of literary texts quantitative data might need to be dealt with differently compared to the analysis of larger amounts of language. This is because corpus stylistics necessitates detailed qualitative analyses to complement quantitative analysis, as the stylistic significance of quantitative data cannot be assessed on the basis of statistical significance alone. This is shown in Stubbs’s (2002) short study of James Joyce’s “Eveline”, where he looks at lexical patterns in the text, focusing on the type-token ratio as a measure of lexical diversity (cf. also O’Halloran (2007) on corpus stylistics and “Eveline”). The patterns of peaks and troughs Stubbs (2002) identifies do not represent a result in themselves, but need to be analysed on the basis of what effect they can have on the reading and interpretation of the text. He shows how the three parts that the peaks and troughs define correspond closely to the way literary critics usually divide the short story. Stubbs (2002: 143) explains that “[w]e must distinguish between the formal features which the computer finds in the text and the human reader’s interpretation of the text, and we must assess the literary significance of what the computer finds”.

In this section, I have discussed some studies that show how the computer can be used to analyse literature, mainly focusing on corpus stylistic approaches to the identification and examination of patterns in literary texts. The aid of the computer makes it easier to spot these patterns, but also provides new ways to study their functions. This of course applies also to the stylistic study of translation. In the next sections, I will introduce translation to the discussion, explaining the role comparison

plays in its analysis and how corpus methods are an effective way to compare different translated texts.

5. Comparing translations

The present research discusses the effects of translating on the link between the linguistic level of HoD, its fictional world, and its major themes. This will be achieved through the comparison of the TTs with the ST, in order to investigate whether discrepancies from the original occur and, if so, what they lead to. It is important to define, then, the basis on which the comparison is established. Do we compare the TTs on the basis of their linguistic form, or do we focus on the maintenance of the meaning? To answer this question, I will look at the relation between form and meaning in translation (Section 5.1), arguing for a functionalist perspective on the way the two interact. In Section 5.2, I will introduce the difference between equivalence and function in translation studies, explaining why comparing the ST and the TTs in terms of functions is more beneficial for the present study. Finally, in Section 5.3, I will explain how I relate this functionalist approach to the study of style in translation.

5.1 Form, meaning, and function in translation

The relation between form and meaning has been at the centre of much linguistic debate. Two distinct perspectives have traditionally dominated the debate. On the one hand, the monistic perspective rejects the form/meaning opposition and assumes that there is no separation between the author's artistic expression and the linguistic means with which it is conveyed. On the other hand, the dualistic perspective accepts that form is one of the possible variants in which a meaning is expressed. In these terms, style is seen as the result of the writer's choices, whereby he/she formulates a given meaning in the way *x*, instead of the way *y* (see Leech & Short 2007 for a discussion of the two perspectives).

If we accept the monistic perspective, it would not make any difference whether we establish the comparison on the basis of the linguistic form or on the maintenance of meaning, because they are inseparable, they are two sides of the same coin. However, if monism were always true, then it would not be possible at all to translate a literary text, as the original meaning would be completely lost in the shift from the source language (SL) linguistic form to the target language (TL) linguistic form. This is counterintuitive, for we know that, although with some loss, translation is indeed possible. Conversely, working within the dualistic perspective, one must accept that it is possible to reproduce the same meaning in several different ways. This would imply that formal alterations in translation may not affect at all the conveyance of meaning: as long as it does not alter the meaning, translation would act as a “neutral reproduction”. We know that this too is not always the case. Therefore, both these perspectives would not be entirely satisfactory to explain what happens to meaning when form is shifted from one language to the other.

An alternative to these opposing approaches is provided by the pluralistic perspective (Leech & Short 2007: 24), which assumes that language performs a variety of functions, and any sample of language is the result of choices made on functional levels. Pluralism maintains that language is mostly multifunctional, so the dichotomy between form and meaning would be restrictive in most of the cases. Jakobson (1960) distinguishes six different functions (conative, referential, emotive, poetic, phatic, and metalinguistic) and explains that the verbal form of a piece of language depends on its predominant function. However, there is always an interaction between predominant and subsidiary functions, as no message has only one function. Similarly, Halliday & Matthiessen (2004) recognise three fundamental language metafunctions (ideational, interpersonal, and textual) that, through their interplay, encompass all aspects of language use. Conceptualising the relation between form and meaning in terms of functions enables me to see stylistic choices as interrelated with functional choices. Linguistic features can be regarded as participating in the enacting of a range of different functions, instead of being tied to an unequivocal meaning. Within the aims of the present study, the TTs comparison can focus on functions across texts, and alterations in the linguistic form can be studied as affecting the reproduction of the original function in translation.

5.2 Function and equivalence

The concept of function is equally significant in translation studies, as it symbolises a specific approach to translation. Venuti (2004) observes that the history of translation theory can be imagined as the changing balance of two categories: equivalence and function.

According to Venuti (2004: 5), ‘equivalence’ (also understood as ‘accuracy’, ‘correctness’, ‘adequacy’, ‘correspondence’, or ‘fidelity’) is a “variable notion of how the translation is connected to the foreign text”. It is this connection, i.e. the equivalence, which makes it possible for the TT to be regarded as a translation of the ST (Kenny 2008: 96) in the first place. Equivalence can refer to the relationship between entire texts, as in the ST-TT equivalence, but it can also indicate a relation between parts of the texts. Catford (1965) considers equivalence as both a theoretical and empirical phenomenon. He adopts a linguistic approach to study the empirical aspect of equivalence, defining it as ‘textual equivalence’. Textual equivalence (Catford 1965: 27) is “any TL text or portion of text which is observed on a particular occasion [...] to be the equivalent of a given SL text or portion of text.” Although this linguistic approach has been criticised (Snell-Hornby 1988), Kenny (2008: 97) recognises that not many alternative ones have been put forward and, especially in the context of translation practice, equivalence is often understood in Catford’s (1965) terms today. One of the alternative views is proposed by Baker (2011), who tries to connect this linguistic approach to an extra-linguistic understanding of the concept of equivalence. Her discussion covers the different textual levels at which equivalence can be established, from one-to-one, word equivalence to grammatical and pragmatic equivalence, but also touches on aspects of equivalence beyond the textual level, for example the ethical and moral implications of linguistic choices.

Whereas equivalence is usually established between the translated text and the foreign text, the notion of function links the translation to the foreign context. Venuti (2004: 6) defines function as “a variable notion of how the translated text is connected to the receiving language and culture”. From a functionalist perspective, translation is perceived as a communication act and understood in terms of the functions it establishes in context (Schäffner 2008: 115). The linguistic form that the TT takes is therefore defined by the purpose, the function, it is expected to fulfil. In this way,

translation acquires a wider potential that goes beyond its association to the ST: from the simple communication of information to the reproduction of a response comparable to that produced by the ST in its own culture. Finally, it is not only the ST that defines the translation process and form, but also the situational and cultural context in which the TT is meant to function (Schäffner 2008: 116). A functionalist concept of translation can therefore encompass a broad range of extra-linguistic factors, so as to offer an explanation when it is impossible to establish relations of equivalence for every translation situation (Venuti 2004: 6). This potential fits the aims of the present study, because the goal of the comparison of the TTs is not limited to the identification of linguistic differences but also tries to account for the effects of such differences on the response to the TTs.

Comparing the Italian TTs on a functional basis brings several advantages to this study. First, it avoids being caught in too strict an understanding of how form relates to meaning. Form and meaning are connected in translation, but their relation is not as absolute as the monistic perspective claims. At the same time, their relation is not as loose as the dualistic perspective assumes. Functions are a more effective way of conceptualising how form conveys meaning. Establishing the translation comparison in functional terms allows me to see alterations in the lexical form as affecting one of the many functions that a given feature enacts. This is in line with Tognini-Bonelli's (2001) approach to the translation practice and her recommendation to establish translational equivalence on a functional basis, rather than on word correspondence. She proposes a procedure that, through a series of steps, moves from formal patterning to function in the SL and then from function to formal patterning in the TL (Tognini-Bonelli 2001: 133). Functions, then, are not necessarily tied to individual words, but rather to the wider contextual patterning associated with them (Tognini-Bonelli 2001: 150). I will return to this procedure in more detail in Chapter 3. Suffice it to say here that the choice of focusing on functional equivalence finds support in Tognini-Bonelli's (2001) approach, which argues for relating formal parameters to functional ones in inter-language analyses.

The second advantage of comparing the Italian TTs on a functional basis is the possibility of extending the concept of function beyond the textual level. Not all of the translators' choices can be studied as inter-linguistic shifts; at times, extra-linguistic factors may be involved. In these cases, it is important to relate observable language

phenomena to non-observable norms and situations (Saldanha 2009: 3) in order to obtain a better understanding of the translational shift. Comparing translations in terms of functional equivalence, that is, looking at how the translation relates to the target language and culture instead of how it relates to the ST, allows me to account for extra-linguistic factors when necessary. In other words, the functional alterations of the TTs can be studied in terms of their effect on the reception of the text in the target culture, rather than in terms of linguistic discrepancies from the formal parameters of the ST only. A similar point is raised by Munday (2008), who adopts a systemic-functional model to analyse the relation between style and ideology in the context of English translations of 20th century Latin American writing. Munday (2008: 31) borrows Halliday & Matthiessen's (2004) functional approach to language analysis to explain how shifts at the lexico-grammatical level can, in certain cases and reiterated throughout the whole translation, shift the higher framework of the text, including the cultural and ideological macro-contexts.

5.3 Function and style

The focus on functions in translation does not imply that the analysis of style moves to the background. On the contrary, it is through the examination of the linguistic level that the present research can compare functions across translations. Functions are in fact enacted through language, so their comparison necessarily involves examining the formal features through which they are established. Stylistics, in Leech & Short's (2007: 13) words, is exactly "the study of language as used in literary texts, with the aim of relating it to its artistic functions".

Given its role in conveying artistic functions, then, preserving the style of the original is fundamental in literary translation, specifically for two interrelated reasons (Boase-Beier 2004: 280): (i) to maintain the same potential range of artistic effects in translation and, at the same time, (ii) to keep an equivalent range of possible interpretations. Any alterations to the style of the ST can therefore affect both the linguistic and extra-linguistic level of the translation, potentially transforming the reading experience:

Style is absolute, you take it or leave it. And if it is lost in translation, then, presumably, the author is lost too, the individual vision is lost; we are left with a text that may or may not be successful, that may or may not be full of interesting ideas, but it is Kundera or Lawrence no longer. (Parks 2007: 240)

Thus, the present study will compare style as carrier of artistic functions across translations. Based on the findings of the stylistic analysis of the ST, the comparison of the TTs will aim to show whether stylistic alterations take place in translation and, if so, what effects they have on the artistic functions conveyed by the original. In this respect, this research can be considered as ST-oriented. An ST-oriented stylistic analysis is generally based on the close comparison of ST and TT, and relies on this comparison to provide answers to the research questions (Saldanha 2014: 101). The focus on the textual comparison may lead to the assumption that any striking stylistic feature in the TT is a reproduction of – or attempt to reproduce – a corresponding striking feature in the ST (Saldanha 2014: 101). However, not all the answers can be found in the ST, as translating often involves other parameters independent of the original: “the mediator’s interpretation of the original; the purpose of the mediation – bearing in mind that the purpose the translation is intended to serve may differ from that of the original; and the audience for the translation” (Saldanha 2014: 101). These parameters take centre stage in the present research, especially the mediator’s interpretation of the original (see Chapter 6). I argue that comparing the TTs in terms of functions, instead of being limited to linguistic equivalence, helps broaden the range of the analysis so as to also encompass extra-textual factors. In this way, marked stylistic features of the TTs do not need to be explained necessarily in terms of the stylistic features of the ST. Alternatively, they can be seen as developing different functions from the ST, functions that may be motivated by the mediator’s agenda. Ultimately, although ST-oriented, the present research aims to explore both the linguistic and non-linguistic aspects that are involved in translation.

6. Corpora and translation

Corpus linguistics and descriptive translation studies, as theorised by Holmes (1972) and Toury (2014), share a number of underlying tenets (Laviosa 2002). First, both approach their objects of study from an empirical perspective and examine real-life language samples, rather than using intuition-based assumptions. Second, both assert that generalisations of empirical findings need to be based on the study of large amount of examples, as opposed to individual instances only. Finally, both approaches aim at describing probabilistic rules of behaviour in the analysis of their object of study, rather than formulating prescriptive assertions (Laviosa 2002: 16). This common ground has enabled a methodological interaction between the two fields and today corpus methods are frequently used to investigate translation-related issues. Similar to the “corpus turn” in stylistics (Leech & Short 2007: 286), corpus approaches to the study of translation have provided both more efficient ways to answer questions in translation studies and suggested new questions altogether. Shifting Tymoczko’s (1998: 1) famous words from 1998 to today, the corpus turn in translation studies helps the discipline to “remain vital and move forward”.

In this section, I will discuss the use of corpus methods to study translation. I will first argue that the interaction between corpora and translation studies has resulted in a wide variety of different approaches (Section 6.1). I will then focus on the specific approach adopted in this thesis, discussing relevant studies in the field (Section 6.2). I will also illustrate more computational methods to the study of translation (Section 6.3), while in Section 6.4 I will discuss the difference between the analysis of the translator’s style and that of the author’s style in translation.

6.1 Corpus approaches to the study of translation

Since its early days, ‘corpus-based translation studies’ (CTS, as defined in Laviosa 2002, Kruger et al. 2011, and Oakes & Ji 2012) has developed into a range of different strands that encompass a wide variety of approaches and aims. For example, corpus methods have been used in contrastive linguistic studies (Granger et al. 2003). In this field, the availability of parallel and comparable corpora has offered

unprecedented perspectives on the similarities and differences of languages in use, enabling inter-language research on the frequency distribution of linguistic features across languages and stylistic preferences (Johansson 2003: 40). An example of this strand is Lavid et al.'s (2009) study of thematisation in English and Spanish in three different sets of data: comparable original texts, translations from English into Spanish, and translations from Spanish into English. This threefold investigation allows them not simply to have a contrastive view of thematic head choices in English and Spanish, but also to compare the effect that translation plays on this linguistic phenomenon in respect to untranslated language.

Corpus methods have also been used in the area of translator training. Baker (1999: 287-288) argues that the use of corpus resources is particularly useful in the cases where translation students are learning in a foreign language environment and cannot rely on their mother tongue intuition to decide whether something sounds natural in the TL or not. In these situations, corpora can provide data to check and compare the equivalence of patterns in the working languages.

Most commonly, corpus methods have been employed to compare translated language to non-translated language, in order to identify the distinctive features that characterise translation as a language variety in its own right. Baker (1993) is usually seen as the forerunner of this approach, which aims to understand “what translation is and how it works” (Baker 1993: 243). Corpora of translated texts are compared to corpora of non-translated texts to emphasise the similarities and differences between them. Baker (2004), for example, examines the frequency and distribution of a range of 3-word, 4-word, and 5-word fixed and semi-fixed phrases across translated and non-translated English texts. This application of CTS has led to considerable research in the area of ‘translation universals’, that is, those linguistic features that typically occur in translated texts as the result of the mediation between two languages, independent of the pair of languages involved (see Baker 1993, Laviosa 2002, Olohan 2004, and Laviosa 2008 for a critical appraisal). Øverås (1998), for instance, uses corpus methods to investigate Blum-Kulka’s (1986) ‘explicitation hypothesis’. According to this hypothesis, the level of cohesive explicitness tends to increase in the TT, regardless of the languages involved in the translation practice. In order to prove this, Øverås (1998) looks at the frequency of additions and specifications as examples of cohesive explicitness in a corpus of 20 Norwegian novels and their English

translations, as well as 20 English novels and their Norwegian translations. Øverås (1998: 16) is able to confirm Blum-Kulka's (1986) hypothesis, providing quantitative evidence to support this.

Grabowski (2013) adopts principal component analysis and cluster analysis to investigate levelling out in a corpus of literary Polish. In particular, multivariate analyses are used to measure the variation between translated and non-translated texts, in order to verify whether translations are more similar to each other compared to original texts and translated texts (Grabowski 2013: 265). His results confirm the levelling out hypothesis, although Grabowski (2013: 275) calls for more quantitative and qualitative research that could replicate the study in more varied language registers or in a larger corpus of translated Polish. A different application to the study of universals is that proposed by Redelinguys & Kruger (2015), who look at explicitation, simplification, and normalisation as an indication of translator expertise. They hypothesise the existence of differences in the occurrence of these phenomena across the writing of inexperienced translators, experienced translators, and non-translated writing. Their findings support the shared conception that differences exist between translated and non-translated language, also providing evidence that the frequency and distribution of universals vary depending on the level of expertise of the translator. However, explanations for these phenomena remain speculative (Redelinguys & Kruger 2015), as there is no means to test why these features appear with different frequency in different texts.

The overall speculative nature of translation universals has often been pointed out, since it is difficult to assess whether these phenomena are actually universal or simply contextual, occurring in language-specific contexts. Ulrych & Anselmi (2008) discuss exactly this, aiming at developing a distinction between language-specific and universal features of mediated discourse. They indeed advocate for this distinction, especially as far as English is concerned: given its role as a lingua franca, English is particularly receptive to hybridism and, as such, varies and changes through usage by non-native speakers and interaction with other languages via translation (Ulrych & Anselmi 2008: 270).

The application of CTS proposed in this thesis is different from the strands discussed so far. My application focuses on the relation between specific STs and their TTs, and examines how text and context are intertwined in translation. This

follows a more and more widespread awareness in the field of CTS, according to which linguistic description and comparison alone are no longer enough (Hermans 1999, in Laviosa 2011: 24). Context-specific phenomena need to be addressed alongside linguistic-specific ones: translation is hardly separable from the socio-cultural background in which it takes place. This application aims therefore to bridge linguistic and cultural studies, socio-cultural and literary investigations, in order to account for the multifaceted nature of the translation phenomenon.

6.2 Corpus stylistics and translation studies

The approach to CTS adopted in this thesis looks beyond the text to contextualise the data. This implies the necessity of combining perspectives and methods from different disciplines (Saldanha 2009: 6), to avoid the emphasis on the text that the use of corpus tools necessarily entails. In this respect, this trend tries to achieve multidisciplinary in a similar way to how corpus stylistics aims at multidisciplinary when analysing literary texts. It is not coincidence that, especially in the ambit of literary translation, this combination of methods comes through the use of a corpus stylistic approach, with its attention to qualitative as well as quantitative analyses and its interest in the critical interpretation of the text.

There is in fact a parallelism between the approach to CTS adopted in this work and corpus stylistics, while what has been described in the previous section can be seen as closer to corpus linguistics. On the one hand, the emphasis on translation universals reflects a corpus-linguistics approach to the study of translation. Instead of examining individual texts, the focus is on translation as a language variety in its own right (Baker 1993). The analysis looks for regularities across whole corpora, in order to develop general theories that can describe the characteristics of the translated language (Øverås 1998, Grabowski 2013, Redelinguys & Kruger 2015). On the other hand, the development of CTS described in this section adopts a corpus stylistics attitude towards the analysis. In this case, individual texts are used, rather than general corpora. Comparison is usually established at the level of ST/TT, instead of non-translated language/translated language, and the resulting findings are interpreted in order to mirror the specific context of production of each translated text. The focus is

on the discrepancies between the original and the translation, and on their effects on the translation reception by the target reader.

This trend is exemplified, for instance, in the work by Bosseaux (2004, 2006) and Winters (2005, 2007, 2010). Bosseaux (2004, 2006) analyses the French translations of Virginia Woolf's *The Waves*², focusing on the rendition of *you* in Bosseaux (2006) and on deixis, modality, and transitivity in Bosseaux (2004). The aim is to show that the alteration of these features in translation not only affects the linguistic surface of the novel, but also its overall framework. In one of the passages examined in Bosseaux (2004), for example, the rendition and repetition of *I* and *must* in relation to the character Bernard vary in the two translations. These changes alter the way Bernard comes across, as the depiction of his mind style is affected by the translators' choices. Bosseaux (2004: 272) is therefore able to use the data provided by the computer to study narratological aspects of the TTs. Likewise, the difference in the translation of *you* (Bosseaux 2006), which in French can be rendered with the informal *tu* (second person singular) or *vous* (second person plural and formal second person singular), affects the formality of some characters and the overall register of the text. As a result, one of the translators renders the novel more formal, while the other conveys more intimacy and familiarity. Again, Bosseaux (2006: 609) shows that "microtextual shifts have consequences on the macrostructure and point of view of the novel".

A similar point is raised by Winters (2005, 2007, 2010), who analyses point of view, speech and thought presentation, and modal particles in two German translations of F. Scott Fitzgerald's *The Beautiful and Damned*. In particular, Winters (2007) focuses on the repetition and variation of speech-act report verbs in two different TTs. She identifies the most frequent speech-act report verbs in the original and then compares the frequency and variation of their translation in the two German versions. Matching the quantitative data with the qualitative analysis of textual extracts, Winters (2007: 424) engages with the relation between translator's choice and adherence to a perceived language norm. She shows how the difference in rendering these verbs can be interpreted in terms of whether the translators conform or not to the convention of avoiding repetition when possible. Another particularly

² Cf. Maczewski (1996) for another corpus-based study of Woolf's *The Waves* in French and German, especially to notice the sensible development of the field in these 10 years.

revealing case is that of the modal particle *wohl* (Winters 2010) because it demonstrates clearly how linguistic alterations affect the overall reading of the text. Modal particles do not exist in English. The addition of the modal particle *whol* by the German translators introduces an element of insecurity to the characters (Winters 2010: 182) which is absent in the original but is nevertheless conveyed by *whol*. Even though this choice is moved by (target) linguistic constraints, Winters (2010) argues that it may reveal something about the translator's attitude towards the text, in terms of his/her own perception of the character as being insecure.

Mahlberg (2007) as well shows that readers' perceptions of characters can be distorted in translation, when the translators fail to reproduce what can initially look like unimportant linguistic features. This is the case of the cluster *his hands [...]* *pockets* in Charles Dickens's *Bleak House* and its German translation. Mahlberg (2007: 130) reports that the translator mainly picks up the physical activity related to the cluster, but the contextualisation of this action within the behaviour of the character is less clear than in the English original. Consequently, the characters in the German text lose part of their characterisation tied to the contextual meaning of putting their hands into their pockets.

Čermáková (2015) too studies clusters in translation, focusing on their repetition as a potential signal of stylistic relevance. Her investigation of John Irving's *A Widow for One Year* and its Czech and Finnish translations looks at the repetition of two 10-word and 12-word clusters in the ST and in the TTs. She finds that the alteration of the verbatim form of the clusters affects the stylistic effect their repetition creates in the original. This is particularly true in the case of the cluster *a sound like someone trying not to make a sound*. This cluster is not only repeated verbatim many times, but it also plays with the repetition of the word *sound*. Moreover, Čermáková (2015) proves that *sound* is also a keyword, related to one of the main themes of the novel. In this way, Čermáková (2015) manages to link cluster and keyword analysis, showing how different linguistic features and patterns are interconnected to create literary effects.

Li et al. (2011) adopt a less quantitative approach, using more basic corpus methods and emphasising instead the extra-linguistic context of production. Li et al. (2011) use type/token ratio and sentence length to compare two English translations of Cao Xueqin's *Honglouloumeng*, relying on a comprehensive apparatus of contextual

information to interpret the statistical data. They look at the translators' mother tongue, their second language, where they lived and studied, the position they held, and their mode of translation (Li et al. 2011: 5), together with the translators' comments on their version and on their translation philosophy in general. In this way, they hypothesise that, for example, one of the translations uses a wider range of vocabulary because its translator opted for a more faithful rendition that aims to "introduce the Chinese literature and culture to the English-speaking world" (Li et al. 2011: 9). As a result, a wide range of words are employed in order to translate literally all the cultural references and allusions. The other translator instead uses longer sentences, probably because of his mother tongue, English, which tends to have much longer sentences than Chinese, which on the contrary displays more paratactic structures (Li et al. 2011: 11).

Kenny (2001: 20) argues that linguistics in translation studies has been traditionally seen as leading only to abstract description of a "prescriptively determined language system". However, her study and the studies discussed below have contributed to countering that view, proving that a (corpus) linguistic approach to translation would not necessarily be oblivious to literary, cultural, or ideological considerations. On the contrary, this trend leads exactly towards the establishment of a connection between 'micro-linguistic events' and 'macro-social structures' (Saldanha 2009: 6), in order to obtain a more nuanced understanding of the relation between text, context, and translation.

6.3 Computational approaches

Computational methods have also been used to study translation, although less frequently compared to their use in the field of stylometry and computational stylistics (see Section 4.1). An example of such an application is McKenna et al.'s (1999) study of Beckett's Trilogy (*Molloy*, *Malone Dies*, and *The Unnameable*) and in its French translation. They use principal component analysis to compare the frequency patterns of the most frequent words in the STs and in their TTs, segmented in 20 sections each. They find that, although the texts' most frequent words change from English to French, it is still possible to identify similarities in the patterns across the sections.

This result is interpreted as an evidence of the maintenance of ‘stylistic overtones’ (McKenna et al. 1999: 169; cf. Chapter 7) across translation, i.e. the reproduction of an equivalent textual style that resists the inter-language shift.

A similar result is obtained by Rybicki (2006) in his comparison of character idiolects in Henryk Sienkiewicz’s Trilogy (*Ogniem i Mieczem*, *Potop*, and *Pan Wolodyjowski*) and in its two English translations. Rybicki (2006) builds on Burrows (1987) and, analysing various character configurations (major characters, recurring characters, characters participating in the public/political/love plot, and female characters), discusses whether equivalent idiolects are preserved in the character dialogues in the TTs. His analysis shows that idiolects are indeed transferred from the STs to the TTs, despite the translational and inter-language shift. Grabowski’s (2013) study, mentioned in Section 6.1 in relation to universals, can also be regarded as an example of computational methods in translation studies. In fact, Grabowski (2013) adopts principal component analysis and cluster analysis to test two translation universals: simplification and levelling out (see Laviosa 2008 for simplification and Baker 1996 for levelling out).

Computational methods have also been used to study translators’ authorship, in the same way as they have been employed to investigate original writers’ authorship, although less commonly. Burrows (2002b), for example, applies his measure of stylistic difference, ‘Delta procedure’ (Burrows 2002a), to the comparison of fifteen English translations of Juvenal’s “Tenth Satire” with corpora of original writings. The goal is to test the reliability of Delta in identifying the relation between the translations and the original texts by the same author. Hung et al. (2010) use principal component analysis and cluster analysis to support a hypothesis about the attribution of 24 Buddhist sutras to the same translator or group of translators, given that they were traditionally considered to be translated by different translators. Analysing the position of the individual sutras on the plots resulting from the multivariate analyses, they are able to establish the relationships between the sutras, based on how closely and densely they cluster together.

Finally, Rybicki & Heydel’s (2013) study focuses on collaborative translation. Here, multivariate analysis is adopted to attribute the specific contribution of each of the two different translators who collaboratively translated in Polish Virginia Woolf’s *Night and Day*. In particular, they apply cluster analysis to Delta-normalised word

frequencies (Burrows 2002a), showing that this method can recognise precisely where the shift between the two translators takes place in the text. When comparing different works by the same translators, Rybicki & Heydel (2013) argue that the original author's signal seems to prevail over that of the translator, i.e. there is a greater likeliness between two translated texts by the same author than between two translated texts by the same translator. For example, the two sections of Woolf's *Night and Day* by different translators are more similar to each other than each translator-specific section with other translations by the same translator. They conclude that this finding opens new possibilities for the study of translation as a textual genre on its own right, where translational agency can be seen as authorial: "stylometry may help to define the 'filter' that shows the translator's multifaceted identity as an artist" (Rybicki & Heydel 2013: 716).

6.4 Translator's style and author's style in translation

Before concluding this section, it is worth mentioning that there is another approach to CTS that, even though it adopts corpus stylistic methods as the present thesis, focuses instead on the style of the translator, as distinguished from that of the original author. Baker (2000) has proposed a methodology to identify and investigate translator style, structured on two key points: (i) the study of translator style is TT-oriented, mainly based on the analysis of TTs without considering the corresponding STs; (ii) translator style is represented by the characteristic features of the translator's language which are not a response to the ST style. In this respect, Baker (2000) sees translators as possessing linguistic "thumb-prints" or "fingerprints" in their own right, independent from the original author's linguistic features. If one wants to demonstrate that translation is not only a reproductive activity, but a creative one as well, than it is important – Baker (2000: 262) claims – that we explore not only authors' but also translators' style.

Baker's (2000) methodology has been followed and developed further by Saldanha (2011a), who offers a revised definition of translator style based on four key points: translator style is (i) recognisable across a range of TTs by the same translator, (ii) distinguishable from that of other translators, (iii) coherent in its patterns of

choice, and (iv) has a discernible function (Saldanha 2011a: 30). Again, the emphasis is mainly on the TT, free from the influence of the original's style. This approach is exemplified in Saldanha (2011b), where the styles of two translators are compared on the basis of their use of foreign words. The two corpora used include the English translations by Margaret Jull Costa and by Peter Bush, respectively. Their comparison reveals different trends in dealing with culture-specific items, which are typical of the translators, rather than responses to ST features, and reflective of the way they see themselves in their role of intercultural mediators (Saldanha 2011b: 257).

Wang & Li (2012) focus instead on Baker's (2000) notion of translator fingerprints, examining two Chinese translations of James Joyce's *Ulysses*. In order to identify typical features of each translation, they use keyword analysis to compare one TT to the other. Specifically, they examine the preferences of each translator for particular verbs and Chinese emotional particles. They also compare the frequency of post-positional adverbial clauses across the two TTs, considered as a rhetorical effect that weakens the importance of the dependant clause. Wang & Li (2012) conclude that the differences between the two translations confirm the existence of translator fingerprints. However, instead of arguing for the independence of the translator's style, they claim that it is a result of both the influence of the translator's first language and of the SL (Wang & Li 2012: 91).

A similar position is taken by Huang & Chu (2014) and Huang (2015). Huang (2015) distinguishes between 'S-type' translator style and 'T-type' translator style: the former is the result of the strategies adopted to cope with specific features of the ST, while the latter refers to the idiosyncrasies of the translators themselves. He theorises a parallel methodology that encompasses the analysis of both S-type and T-type stylistic features (Huang 2015: 54), in order not to ignore entirely the role that the ST plays in the translation process. This methodology is shown in Huang & Chu (2014), where both S-type and T-type perspectives are considered. In this study, Huang & Chu (2014) analyse a corpus of English translations of contemporary Chinese novels. They identify a characteristic feature of one of the translators, i.e. the preference for the use of indirect speech with the third person and the past tense to deal with the Chinese omission of personal pronoun subjects and lack of tense markers. They suggest verifying whether this S-type feature (a strategy used to cope with an aspect of the ST) occurs consistently in all the translations by the same translator, so as to

confirm whether it is also a T-type feature, that is, an idiosyncrasy of the translator him/herself (Huang & Chu 2014: 138).

The approach to the relation between ST and TT adopted by these studies is more TT-oriented compared to that adopted in the present study. I argue that this research is mainly ST-oriented. In this respect, the features of the TTs identified are analysed in terms of their relation to ST features, rather than as characteristic of the translators' style. The point of departure will always be the style of the original author, so the TTs are examined in order to see to what extent the author's style is maintained, and what effects discrepancies in style produce on the reception of the TT.

7. Investigating manipulation

Translation is not simply a mechanical transposition of a text from one language to the other, but rather a complex social phenomenon. Translating is so imbued with cultural, ideological, and political implications that it is impossible to detach it from the context(s) in which it takes place. Translators are the mediators between these contexts, and as such they can act as "crucial agents for social change" (Tymoczko 2010: 3). Their agency is reflected in the choices they make, or even in the choices they do not make. This decision-making power can have ideological implications and some translators have used it to perform acts of resistance and activism: supporting language movements, shifting cultural values, gender struggles, etc. (Tymoczko 2010: 229). In the specific context of literature, translation has been equally seen as playing "a most active" role (Even-Zohar 1990: 46) in any literary system. Translating literature can establish affiliation and construct identities, challenge societal norms as well as reinforcing them, and contribute to literary innovation or fixation alike (Even-Zohar 1990). In the final analysis, translation influences the context in which it is produced and, at the same time, is influenced by it.

According to Lefevere (1992), the influential role of translation in literary systems is due to its potential to substitute the original: "[translation] is able to project the image of an author and/or a (series of) work(s) in another culture" (Lefevere 1992: 9), replacing the ST for those who have no knowledge of the SL. This projection,

‘rewriting’ in Lefevere’s (1992) own words, is far from a neutral reproduction: “[a]ll rewritings, whatever their intention, reflect a certain ideology and a poetics and as such manipulate literature to function in a given society in a given way” (Lefevere 1992: vii). There are two mechanisms in control of this manipulative process, one operating outside the literary system, the other inside of it. From the outside, the process is controlled by the ‘patronage’ (Lefevere 1992: 11), i.e. those powers – both individuals and institutions – which “can further or hinder the reading, writing, and rewriting of literature” (Lefevere 1992: 15). Sovereigns, presidents, royal courts, social classes, religious institutions, political parties, publishers, and the media are all examples of patronage. From the inside, the manipulative process is put into practice by the ‘professionals’ (Lefevere 1992: 14), e.g. critics, reviewers, translators, teachers, scholars, and anyone who works directly with literature and its products. Being mainly concerned with the ideological side of literature, patronage acts through the professionals to ensure that this aspect is aligned with its agenda. The professionals rewrite works of literature to make sure that their ideology is deemed acceptable to the norms and conventions of the context of production. Through this system, patronage and professionals can manipulate translation and its reception in the target culture. These ideas underpin what in translation studies is generally referred to as the ‘manipulation school’ (Hermans 1999: 8), a paradigm which assumes that “[f]rom the point of view of the target literature, all translation implies a degree of manipulation of the source text for a certain purpose” (Hermans 2014: 11).

The present study looks at textual discrepancies between the ST and the TTs. Although ideological agendas cannot be comprehensively pictured by linguistic alterations alone, the analysis of stylistic choices in the TTs can reveal the translators’ manipulative presence (Munday 2008: 14). In line with what has been shown in recent studies (cf. Section 6.2), I argue that changes at the linguistic level of a text can affect its interpretational level and that these changes can be motivated by the translator’s agenda. A stylistic comparison of ST and TT is therefore a suitable method to emphasise potential manipulations, in cases in which the choice of the translator cannot be explained on the basis of ST features alone. But how do we recognise a choice that is not motivated by features of the ST as manipulative?

The answer to this question rests on the definition of manipulation. However, despite the centrality of the issue in the discipline, translation manipulation remains

an “underexplored phenomenon” and, as a result, lacks a “comprehensive and unequivocal definition” (Dukāte 2009: 15-16). More precisely, manipulation has been discussed in so many different terms that it is difficult to pinpoint the concept precisely. For instance, Venuti (2008: 1) talks about the “translator’s own manipulation of the translating language” in reference to the domesticating strategy dominant in the Anglo-American publishing system aimed at producing TTs that read as fluently as possible. Nonetheless, there is no definition of manipulation and it is not clear whether all the linguistic choices of the translator are manipulative or not. Other scholars provide examples of manipulative phenomena in real translational contexts, examining them either stylistically (Munday 2008) or socio-culturally (Tymoczko 2010). But most of these examples occur in circumstances in which the balance between the two cultures/languages involved is clearly uneven, e.g. in postcolonial contexts. One may ask whether manipulation also occurs in more ordinary contexts, especially if we consider that the theorists of manipulation (Hermans 2014, Lefevere 1992) claim that all translations imply a certain degree of manipulation.

Dukāte (2009) has attempted a systematisation of the phenomenon, aiming to categorise the different types of manipulation discussed in the literature. She distinguishes two main different types of manipulation: ‘text-external’ (Dukāte 2009: 90) and ‘text-internal’ (Dukāte 2009: 101) manipulation. The former is that manipulation that proceeds outside the text. For example, the choice of what not to translate or what to omit completely from the TT can be a form of text-external manipulation. The latter includes all forms of manipulation contained in the text, that is, what the literature in translation studies usually refers to with the term ‘manipulation’. Both of these can be further divided into manipulation as improvement, manipulation as handling, and manipulation as distortion (Dukāte 2009: 89). Finally, all these types of manipulation can either be conscious or unconscious. However, despite the meticulous classification, Dukāte (2009: 112) claims that “everything anybody does or does not do is manipulation”. In other words, her answer on whether translation is manipulation is that it “depends on the evaluator’s vantage point and his/her understanding of the manipulation and expectations of translation” (Dukāte 2009: 130). Thus, it seems that manipulation is a concept that resists constrictive definitions.

In spite of its indeterminacy, it is possible to recognise some characteristics that are usually associated with a shared understanding of manipulation, specifically as it is conceived by the ‘manipulation school’ (Hermans 1999: 8). This paradigm is mainly concerned with studying manipulation in the context of literary translation, as opposed to other types of translation. Literature is more prone to being manipulated and to manipulate because of its cultural and ideological implications (Dukāte 2009: 151). Within the context of literature, manipulation is seen as affecting all instances of translation (Hermans 2014: 11), although to different degrees. Yet, regardless of the extent, manipulation is mostly considered a negative phenomenon, rather than a positive one: it is a distortion of the ST’s truth, which is misrepresented (Dukāte 2009: 75), rather than an improvement in the TT. Most importantly, manipulation is considered to be motivated: there is always a specific purpose behind it. Usually, this motivation is due to cultural, political, or ideological considerations (Dukāte 2009: 46). Translation can be inspired by ideology, or constrained by it, depending whether the translation is in agreement with or in opposition to the dominant ideology (Lefevere 1992: 8). Nevertheless, according to the ‘manipulation school’, ideology always takes centre stage in manipulative phenomena.

These general characteristics will act as guidelines to determine whether a TT stylistic choice that does not seem to be motivated by ST features can be considered as manipulation or not. Although they are not representative of the whole phenomenon, these guidelines reflect a widespread concept in translation studies. As such, their use allows me to test the outcomes of the ST/TT comparisons on the basis of a generally accepted notion in the discipline. Moreover, despite its overall recognition as a matter of fact in translation, manipulation has rarely been investigated from a linguistic point of view. Yet, linguistic approaches have been successfully applied to the study of ideology, so they can be similarly applied to the investigation of manipulative agendas in translation.

Corpus methods in particular have proven useful in examining the relation between language and ideology. Generally speaking, corpus linguistics is seen as being capable of compensating the shortcomings of discourse analysis, providing “a general ‘pattern map’ of the data” (Baker et al. 2008: 295) to facilitate the identification of specific features worthy of further attention, without being arbitrary in the selection. At the same time, corpus linguistics can offer a quantitative

dimension to phenomena that have already been recognised as ideological in the text. For example, the study of metaphors can be supported by the frequency and distribution of lexical patterns associated with them (Baker et al. 2008: 296), while keyword analysis and their grouping into categories can help identify discursive constructions and representations (Gabrielatos & Baker 2008: 10). Moreover, collocations and semantic prosodies can signal the stance of the writer towards a certain topic, making argumentations and ideological stances explicit (Gabrielatos & Baker 2008: 14).

Orpin (2005) shows the benefits of using corpus methods to investigate ideology in language through her investigation of terms related to the idea of “corruption” (*bribery, corruption, cronyism, graft, impropriety/ies, malpractice(s), nepotism, and sleaze*) in the newspaper sub-sections of the Bank of English corpus. Analysing their frequency, distribution, collocational and semantic profiles, Orpin (2005) is able to delineate some linguistic tendencies that can be indicative of general ideological stances. For instance, she points out that “corruption” words with greater negative associations are employed to refer to events abroad (Orpin 2005: 58), while less negative synonyms (such as *sleaze* or *impropriety*) are rarely selected when e.g. Italy is written about. MacDonald & Hunter’s (2013) analysis centres instead on the discursive constructions surrounding the security operations for the 2012 London Olympic Games. Using keyword analysis and in-depth study of collocations and concordance lines, the authors identify four key features of the discourse: exceptionalism, exclusion, prediction, and pedagogization. Linking quantitative analysis and qualitative examination of textual extracts, MacDonald & Hunter (2013) point out how these key features and their linguistic instantiations contribute to establish a precise ideology: the legitimation and implementation of massive security operations, even when these operations affect the privacy of people.

In accordance with the analysis of ideology and discourse in language with corpus methods, the present study aims to provide a linguistic perspective of the manipulation phenomenon. In particular, I will investigate whether the discrepancies identified in the Italian translations reflect the general conception of manipulation.

8. Conclusion

This chapter has traced a discussion that has connected stylistics to literary criticism to translation. The common denominator of this discussion has been the corpus approach. Being comparative in nature, stylistics benefits greatly from the adoption of corpus methods, which expand the stylistician's toolkit and enable new perspectives on the study of literary texts. This does not mean that corpus stylistics focuses only on the description of linguistic features. On the contrary, it uses innovations from corpus linguistics to trigger a more comprehensive understanding of literary texts. The study of translation too is essentially comparative, as it is based on the comparison of an ST to a TT. Translation studies has equally found in the use of the computer a way to enhance empiricism and objectivity, which are at the heart of the descriptive approach to the discipline. But despite the emphasis on the linguistic level that this method entails, the present research follows a line of enquiry for which the extra-linguistic aspects of translation take centre stage in the analysis. A linguistic approach to the study of literary translation is not oblivious to the context of production or to the agenda of the translator, because textual and contextual levels are strictly related in translation. Alterations in the lexical level can have effects on the interpretation of the text, while TT stylistic choices not motivated by ST features can indicate the translator's agenda. Finally, through the corpus-assisted analysis of HoD and its Italian translations, the present study will make a contribution to the methodological dialogue between stylistics, corpus linguistics, and translation studies.

CHAPTER TWO

Africa and Africans in Joseph Conrad's *Heart of Darkness*

1. Introduction

In literary texts, linguistic patterns can both produce stylistic effects and function as triggers for literary themes. Wales (2011) maintains that the goal of stylistics is that of analysing both functions. Stylistics should show how a text works “not simply to describe the formal features of texts for their own sake, but in order to show their functional significance for the interpretation of the text; or in order to relate literary effects or themes to linguistic ‘triggers’ where these are felt to be relevant” (Wales 2011: 400). However, whilst the relation between patterns and artistic effect is a basic assumption in stylistics, the relation between patterns and literary themes is a less explored phenomenon. The present thesis will be equally concerned with both of these aspects. I will study lexico-semantic patterns as formal constituents of Conrad's style and as triggers of fictional themes and the fictional world in HoD.

For this analysis, the role of literary criticism is particularly significant, as it can indicate established and widespread perspectives on the themes of HoD. A review of the reception of the text will bring forward the themes that are generally considered “major” by literary critics, foregrounding aspects of the text to focus on to analyse such themes. Mahlberg (2007a) explains that there are two main ways to select items for closer analysis in corpus stylistics: either taking suggestions from literary insights and arguments from literary criticism or checking the frequency information of the words in the text and picking those whose frequencies appear noteworthy (Mahlberg 2007a: 22). This study opts for the first way, because a frequency-based selection dismisses low-frequency items that can nevertheless play an important role in the process of theme construction (cf. Chapter 3, Section 4.2.1). Themes are more likely to emerge from the interaction of different linguistic features, as the result of the cumulative effect of several items which may or may not occur very frequently. Because the frequency of these items might not be very high, the selection of what to study based on frequency cut-off points would ignore their contribution. Conversely, suggestions from literary insights would not be tied to formal criteria, but would

instead account for the way the text is read and interpreted by critics. Moreover, taking suggestions from literary criticism allows me to discuss aspects of the short novel that are central in the scholarly debate, in accordance with the purpose of making a significant contribution to the critical study of HoD.

Thus, the purpose of this chapter is that of providing an overview of the reception of HoD in order to emphasise those themes that have received more critical attention. In other words, I will “take suggestions” (cf. Mahlberg 2007a: 22) from literary criticism and use them as a starting point for analysing the relation between patterns and themes. Two major themes will emerge from this discussion: “Africa and its representation” and “race and racism”. I will argue that the postcolonial approach to Conrad has emphasised the importance of the imperialist context in the interpretation of his novels. In the specific case of HoD, this emphasis has given new prominence to the way Africa and Africans are depicted, as their fictional representation has been seen to reflect ideologies and attitudes. The following section will introduce Conrad’s canonical position in English literature, while Section 3 will discuss the postcolonial approach to Conrad criticism as one of the most dominant approaches to his texts. The themes “Africa and its representation” and “race and racism” will be discussed in Section 4 and Section 5 respectively. Finally, Section 6 will offer some concluding remarks.

2. Conrad and the literary canon

Since the publication of his first novels, Conrad has been the object of widespread literary attention. At the beginning of his writing career, his work was already highly considered and, despite some criticisms, critics generally acknowledge that Conrad is going to leave a mark in the history of literature. In the first full-length study dedicated to his work, Curle (1914: 1) claims that “Conrad’s work actually does mark a new epoch”. A few years later, Waugh (1919: 267) foresees that “Conrad’s reputation appears to be absolutely assured”. Early reviewers of his books share the same positive considerations (cf. Simmons 2009 or Peters 2008), recognising in Conrad an innovative strength and an originality in style and themes that distinguish him from his contemporaries. This overall approbation continued to grow after his

death (1924) and the definitive consolidation of his literary fame came in the 1940s. Leavis's ([1948] 1973) *The Great Tradition* is widely regarded as establishing Conrad's work in the canon. Leavis (1973: 1) includes Conrad among the greatest English novelists, together with Jane Austen, George Eliot, and Henry James, and, by doing so he helps foreground Conrad as a writer worthy of scholarly attention (Niland 2009). This strong interest in Conrad matches the increasing professionalisation of academic criticism of the 1940s. This process underpins the re-evaluation of Conrad's work and resulted in the publication of numerous serious scholarly studies that finally established Conrad's "modern" reputation (Knowles 2009). Since then, Conrad's work has been firmly located within the literary canon and today critical interest in his novels is livelier than ever.

In more than a century of uninterrupted criticism, Conrad's oeuvre has been discussed in many different ways and from diverse perspectives. According to Ciompi (2005), Conrad has been read from a political, psychoanalytical, semiotic, deconstructionist, metafictional, and imperialist point of view, among the others. Thus, although his position in the canon has remained unaltered throughout recent decades, critics have focused on different aspects of his work, and have interpreted it differently, depending on the historical and socio-cultural period in which they have been writing. In this wide array of differing paradigms, one of the most important and influential perspectives on Conrad's narrative has been the postcolonial approach.

3. The postcolonial approach in Conrad criticism

An important moment in the history of Conrad criticism is the aftermath of the Second World War. Before that, during the inter-war years, Conrad's literary fame goes through a period of "faltering popularity" (Knowles 2009: 68), in which his work is perceived as "out of touch" (Knowles 2009: 68) with the new generation of readers, distant from the emerging concerns of the time. According to Knowles (2009: 68), for instance, Conrad "exclude[s] modern women (and women readers); he depreciate[s] intellect in a scientific age; and he ignore[s] complicated sexual involvements and social reforms". After the Second World War, however, Conrad's work is increasingly seen to align with the current generational mood. He is considered

modern and prophetic for having foreseen the catastrophic effects of supremacist policies. The analysis of his work assists in thinking critically about the destructive results of the extremist national ideologies which were manifested in the war. In particular, it is Conrad's multi-language, multi-cultural background – his being a writer who transcends national boundaries – that makes his work seem suitable for theoretical readings aimed at criticising aggressive ideologies (Niland 2009: 78).

This new interpretative trend reaches its apex in the period that witnesses the breaking apart of the British Empire. In this context, Conrad's involvement with the imperial and colonial question acquires a new meaning. In the early phases of his reception, this involvement was mostly seen as a secondary aspect of Conrad's narrative: it was mainly the setting of his novels that contributed to the exotic feeling that Conrad's work was considered to have. Conrad was in fact often paralleled to Robert Louis Stevenson because of his Malaysian and seafaring settings (Simmons 2009: 60). With the rise of postcolonial perspectives, however, Conrad's engagement with imperialist issues takes centre stage in any interpretation of his work. Said (1994) argues that Conrad's narrative is so embedded in the imperialist ideologies of its time that it would be impossible to disentangle the two without missing Conrad's argument: “[w]ithout Empire, [...], there is no European novel as we know it” (Said 1994: 69). The postcolonial approach has thus established itself as one of the most widespread and discussed approaches in Conrad criticism today.

Edward Said plays a fundamental role in establishing this postcolonial perspective, being a forerunner in applying this approach to literature and, of particular interest to this thesis, to Conrad's writing (Niland 2009: 79). Said's (1994) interest in Conrad is focused on issues of imperialism and postcolonialism. The study of Conrad is simply unthinkable without considering these questions. In Said's (2009) view, Conrad's work can be read as both imperialist and anti-imperialist at the same time. On the one hand, Said (1994) recognises an awareness of the disastrous consequences of cultural imposition in Conrad's work; on the other, an inability to conceive alternatives to this imposition, a different course for this process. Although progressive in some respects, Conrad's work is nevertheless embedded in the ideologies of its own time. As a result, Conrad overlooks the fact that, under the surface of what seemed to him exoticism, India, Africa, and South America had their own independent histories and cultures, which the imperialists consequently trampled

(Said 1994: xvii). Said (1994) finds this perspective perfectly reflected in Marlow and Kurtz:

They (and of course Conrad) are ahead of their time in understanding that what they call “the darkness” has an autonomy of its own, and can reinvade and reclaim what imperialism had taken for *its* own. But Marlow and Kurtz are also creatures of their time and cannot take the next step, which would be to recognize that what they saw, disablingly and disparagingly, as a non-European “darkness” was in fact a non-European world *resisting* imperialism so as one day to regain sovereignty and independence, and not, as Conrad reductively says, to reestablish the darkness. (Said 1994: 30, emphasis in the original)

In the end, Said (1994) suggests that there is no real alternative to imperialism in Conrad. Conrad’s acute critique of the imperialist enterprise condemns its atrocities, but it does not go as far as to conclude that imperialism itself has to cease so that the colonised could live their lives free from European dominance: “[a]s a creature of his time, Conrad could not grant the natives their freedom, despite his severe critique of the imperialism that enslaved them” (Said 1994: 30).

Another major contribution to postcolonial readings comes from the Nigerian novelist and critic Chinua Achebe. Achebe is a particularly influential figure because, by accusing Conrad of being a “thoroughgoing racist” (Achebe 1990: 11) during a 1975 public lecture at the University of Massachusetts, he generates a persistent debate that has further consolidated the importance of the colonial and imperialist context of Conrad’s writings. In contrast to Said (1994), Achebe (1990) argues that Conrad implicitly and explicitly conforms to the imperialistic ideology, and that this is evident in the way he depicts Africa and Africans in HoD. Africa is seen as a “place of negation” (Achebe 1990: 3), the antithesis of what Europe should symbolise, the wilderness in its natural form, without control or ties. It is exactly what contemporary readers expect and Conrad simply acts as “purveyor of comforting myths” (Achebe 1990: 5), giving them precisely what they want to be shown. Similarly, Africans are wild and unbound, so distant from the Europeans that they seem almost inhuman.

Achebe (1990) in fact claims that Conrad conceals every aspect of their humanity, depriving them of language and identity.

Achebe (1990) is aware of the fact that the attitude towards Africa and Africans in HoD could be Marlow's, rather than Conrad's. He recognises the different layers of narration in the text, and the distinction between the author (Conrad), the first narrator (the unnamed man on the *Nellie*), and Marlow (whose account is reported by the first narrator). However, Achebe (1990: 9) sees these literary figures merely as "layers of insulation", only as a narrative artifice. The text does not provide any alternative framework of reference for the reader to judge Marlow's actions and opinions differently from Conrad's (Achebe 1990: 10). Conrad would have been able to do so, if he had considered it necessary, but he did not. Achebe (1990) therefore concludes that Conrad seems to approve of Marlow, also given the similarities between the fictional character's and the author's careers (Achebe 1990: 10). Conrad's responsibility for Marlow's attitude leads Achebe to believe that it was Conrad himself who had a problem with black people, an intimate and maybe unconscious antipathy (Achebe 1990: 13). Even admitting the influence of the dominating prejudices of the time, Achebe (1990) still recognises the attitude towards black people in the text as Conrad's. For example, he attributes to Conrad an "inordinate love" (Achebe 1990: 13) for the word *nigger* and an obsession with blackness.

Despite the wide range of approaches from which HoD has been studied, Achebe (1990) complains that the assumed racism of the short novel has been ignored. Ultimately, he wonders "whether a novel which celebrates this dehumanization, which depersonalizes a portion of the human race, can be called a great work of art. [His] answer is: No, it cannot" (Achebe 1990: 12).

The postcolonial approach to Conrad criticism is relevant to the present study for several reasons. First of all, because this reading of Conrad's work is one of the most diffused and discussed today. Interpretations such as Said's (1994) and Achebe's (1990) have shaped the critical reception of HoD so deeply that it is almost impossible to ignore postcolonial and imperialist issues when studying the short novel. This means that focusing on these aspects will allow my research to contribute to current critical debates. Moreover, given its enormous influence on contemporary criticism, the focus on postcolonial issues will enable me to examine the effect of dominant

readings on translation practice. Said (1994) argues that Conrad's texts can be read both as espousing and opposing the imperialist enterprise, whereas Achebe's (1990) controversial accusation has split the critics between those who agree with him and those who do not. Thus, the issue of racism is clearly debatable and where the translator stands in this debate can be reflected in his/her translation. Even-Zohar (1990: 51) defines translation as "an activity dependant on the relations within a certain cultural system", whose shape is not defined once and for all by its ST. Different translations of the same original can therefore differ depending on how the translation/translator relates to his/her cultural system. According to Lefevere (2014: 238), this relation is inescapable, as no translator can ever escape from the ideology/poetics dominant in the target context of production. It is always possible to see a given translation as belonging to a given socio-cultural context. This analysis will then look for the influence of the postcolonial reading in the Italian translations of HoD, investigating the extent to which this is visible on the linguistic level of the TTs. This is in line with Lefevere's (2014) understanding of the study of literature, which should:

explain how both the writing and rewriting of literature are subject to certain constraints, and how the interaction of writing and rewriting is ultimately responsible, not just for the canonization of specific authors or specific works and the rejection of others, but also for the evolution of a given literature. (Lefevere 2014: 219)

In other words, this will shed some light on the Italian reception of HoD, as it is filtered by the mediating role of the translators. Finally, the fact that the postcolonial approach has emerged in the second half of the twentieth century means that the early Italian translation of HoD has been produced in a period in which these questions are not at the forefront of critical attention. This will provide a more diachronic perspective, as it will be possible to compare the early TT with the contemporary TTs and to study whether the latter have been influenced by their context of production in a way that the former could not have been.

As I mentioned in Section 1, this thesis will focus on two themes that have emerged as particularly prominent from a postcolonial approach to HoD: "Africa and

its representation” and “race and racism”. I will discuss each of these themes in Section 4 and Section 5 below, respectively.

4. Africa and its representation

The “where” of literature has recently been playing an increasingly central role in literary criticism, especially in the context of modernism (Thacker 2005, Brooker & Thacker 2005). The relationship between literature and geography is bidirectional: literature contributes to the understanding of social, geographical places through fictional representation; at the same time, social spaces help to shape literary forms and styles. Thacker (2005: 63) terms the outcome of this interaction ‘textual space’ – the resulting interplay “between spatial forms and social space in written text”. The study of textual spaces goes beyond describing how a given place is represented in the text, but accounts also for how this depiction is culturally and socially constructed. Ultimately, textual spaces connect fictional to real places, but also socio-cultural questions to literary forms and styles. As a result of this interaction, textual spaces have the power to shape the perception of spaces and places both inside and outside literature.

Jarosz (1992: 106) states that, through the use of metaphors and symbolism, the written representation of places have the same potential, although figurative, to construct, shape, and destroy spaces in the same way as historical, material, social, and ecological forces have. Africa has been persistently associated with a series of metaphors; one of the most dominant is Africa as the “dark continent”. Jarosz (1992) argues that this enduring metaphor has played a significant role in defining the perception of the continent. The “dark continent” metaphor has not only influenced how literature, travel accounts, news reports, and academic writing represent Africa, but it has also affected, in a derogative manner, the socio-cultural perception of Africa itself. The persistence of this metaphor has moved Africa’s geographical and cultural variety into the background, putting in its place a homogenising “other”. Africa acts as a blank space, deprived of differences, and functioning as a mere term of comparison for Western notions and values. The perpetuation of this symbolism in the

last hundred years has confirmed and even legitimated domination discourses such as imperialism, sexism, and racism (Jarosz 1992).

Likewise, Hegglund (2005: 44) identifies two other persistent and related metaphors with which Africa has been depicted in the twentieth century: as a realm of timeless myths and as part of the historical process of globalisation. In line with Jarosz (1992), Hegglund (2005) argues that the mythical 'image-Africa' has triumphed over the historically contextualised and real Africa. The image-Africa has homogenised the real Africa, substituting its varied culture with a timeless place out of history which keeps resisting modernity and civilisation. Given this assumed resistance to modernity and civilisation, Africa is also seen as the place where globalisation should take place. The continent therefore emerges as the object of study for explorers and colonists, who impose a Western filter on all non-Western places and things. The effect of this dual metaphor is, again, uprooting Africa from its socio-cultural and historical background, and replacing it with an ahistorical and ageographical image, an empty container into which any number of symbols can be projected (Hegglung 2005: 48).

Within the aims of the postcolonial approach, the way Conrad depicts Africa plays a major role in the interpretation of his novels. Conrad's textual spaces are seen to be in dialogue with issues of power and domination typical of the twentieth century, and his fictional representation of Africa is studied as a reflection of ideologies and attitudes towards both the continent and the colonial enterprise in general. GoGwilt (2005) recognises the key role of Africa and its representation in HoD, arguing that the way the continent is formulated is "a touchstone for the modernist rupture" with the enlightened European history and geography (GoGwilt 2005: 66). Hampson (2005: 56) proposes a contrastive approach that compares Congo in HoD with Congo described in Conrad's personal records of his journey in the African continent. Hampson (2005) emphasises the discrepancies between the fictional representation of the journey along the Congo River in the short novel and the logs reported in Conrad's records, and uses these differences to foreground the author's intention to depict Africa in a metaphorical way.

Overall, since its publication, HoD has helped shape the representation and reception of Africa, by virtue of its fame and critical acclaim. This is true not only in literary contexts but also in wider popular culture. A postcolonial approach to HoD

foregrounds the relation between the text and the imperialist context, emphasising the way Africa is fictionally constructed. Analysing the textual representation of Africa in HoD allows me to establish a link between linguistic features of the text and one of the major themes I seek to discuss: “Africa and its representation”.

5. Race and racism

By accusing HoD of racism, Achebe (1990) has had an enormous impact on academic criticism. His reading has “significantly altered the landscape of Conrad scholarship” (Pursell 2009: 86), challenging the canonised status that, more than 25 years before, Leavis (1948) had granted Conrad. Engaging with HoD today necessarily demands an engagement with what Achebe (1990) has said about it (Allington 2006: 132), whereas before his public lecture the issue was virtually non-existent.

From HoD’s early reception right through to the 1950s and 1960s, there were barely any references to racial issues in Conrad’s short novel. On the contrary, Conrad was praised for “neither dehumanis[ing] nor europeanis[ing]” (Curle 1914: 122) his non-European characters, which are portrayed with “curious fidelity and insight”, and placed in a “world of their own, hidden from our understanding – the world of savage fears and beliefs” (Curle 1914: 122). Referring to HoD specifically, Curle (1914) appreciates Conrad’s representation of “the whole sadness and dark unrest of savage minds – I mean the minds of real, untutored savages” (Curle 1914: 122). Similar responses to Conrad’s work can be found in a number of contemporaneous reviews, criticisms, advertisements, and biographies (cf. Peters 2008), none of which make any reference to racial issues. Randall (1925, in Peters 2008), for example, praises Conrad’s ability to portray faithfully different races and nationalities, without any kind of subjective imbalance: “to those who know them, the Malays, Chinese, Indians, Negroes and Arabs are as convincing as the Englishmen” (Randall 1925: 115, in Peters 2008).

In comparison to the early reader, who would not have encountered a discussion of racism in relation to HoD, today’s readers cannot help but engage with ideas of race and racism, for almost every critical edition of the novel deals with these issues, or at least mentions them. As Allington (2006: 132) observes, readers are

“obliged, in reading *Heart of Darkness*, to take a stand on a matter of controversy that, in the Sixties, simply did not exist – namely, whether or not it is a racist book”. Likewise, today’s critics cannot ignore this controversy. In fact, critical debates surrounding HoD and racism have been ongoing for the past four decades, and have generated a vast amount of research on the topic.

Within these debates, few critics have shared Achebe’s (1990) accusations as directly and passionately as him. Almost every discussion of HoD that has preceded Achebe’s (1990) lecture, however, has acknowledged the pertinence of the issue of racism. Brantlinger (1985), for example, recognises the powerful criticism of imperialism and racism that HoD offers but, at the same time, admits that such a criticism is “characterized as both imperialist and racist” (Brantlinger 1985: 365). Lawtoo (2012) adopts a similar position. He admits that “‘racism’, as Achebe recognised, is clearly part of [Marlow’s] rhetoric” (Lawtoo 2012: 249). At the same time, he suggests that perhaps Marlow is simply attempting to establish a connection with his white listeners through the use of the same racist rhetoric that dominated Conrad’s time. In other words, Marlow is merely providing his listeners with what they expect. Coats (2014: 645) recognises in the use of scientific racist discourse (such as the reference to phrenology in HoD) one of the aspects that most disturbs Conrad’s readers. Although scientific racist discourse is often accompanied by its critique, its inclusion can nevertheless signal “more than a set of passing references to contemporary cultural life” (Coats 2014: 645). Miller (2012), too, recognises the use of racist stereotypes and clichés in HoD:

I want to believe, as many other critics do believe, that Conrad was not a racist and that he was in *Heart of Darkness* attacking the racist side of imperialism, partly by embodying it ironically in Marlow. Nevertheless, the novel employs many racist stereotypes as well as racist clichés from journalism and popular literature of the time. (Miller 2012: 30)

Miller’s (2012) quotation is a relevant one, because it suggests that the theme of race and racism is taken into account even when the critic does not want to agree with Achebe (1990). Zins (1982), for instance, claims that Conrad was an anti-imperialist

and anti-racist precursor, and that HoD manifests his progressive thoughts. Zins (1982: 121) reads the short novel as a direct denunciation of colonialism in Africa, and argues that Achebe (1990) is “definitely doing injustice to Conrad” (Zins 1982: 122) by accusing him of being racist and directing antipathy towards black people. But, even though “in its deepest sense, contrary to Achebe’s opinion, the story is antiracist” (Zins 1982: 147), Zins (1982) concedes that Conrad did not entirely escape the ideological conventions of the imperialist tradition, and uses them to stereotype the Africans (Zins 1982: 125). A similar concession is put forward by Watts (1990). He is very keen to emphasise the importance of the socio-cultural context in the interpretative process, claiming that subordinating HoD to present-day value systems and criticising the text for failing to encompass them may be itself an exercise of ideological and temporal imperialism (Watts 1990: xxvi). According to Watts (1990), HoD can be seen as a progressive and anti-imperialist text when framed in its native context. Nevertheless, some of Marlow’s attitudes towards the Africans remain patronising and misguided (Watts 1990: xxi).

Ultimately, then, Achebe (1990) has not deprived HoD of its canonical status, or completely won over public and critical opinion, but he has been successful in making the academic community acutely aware of potential issues of race and racism in the text. This lively debate has consequentially drawn great attention to the way in which the African natives are represented in the text. Most – if not all – of the critics discussed in this section make reference to how Africans are depicted in order to support or discredit the hypothesis that HoD is a racist text. I will therefore analyse the fictional representation of the African natives as a way to link the linguistic level of the text to this major theme of race and racism.

6. Conclusion

In this chapter, I have looked at Conrad’s established position in the literary canon and have argued that, although his canonical status remains unchanged, critics have prioritised different aspects in their reading, depending on the historical and socio-cultural context in which their interpretations took place. Aspects such as the representation of Africa and of Africans were of little interest to early readers of HoD,

whilst today they are dominant, given the development of postcolonial approaches to literature. The postcolonial approach gives great importance to the imperialist context in Conrad's novels, stressing the relevance of its representation as a reflection of ideological stances and attitudes. This discussion has therefore led to the identification of two major themes in HoD, which are central in critical debates today: "Africa and its representation" and "race and racism". Given their centrality, this thesis will focus on each of these themes in detail.

There are several reasons for why focusing on these themes is a well-grounded choice for the present analysis. First of all, I have argued that they represent central themes in the current critical debate. They not only emerge from one of the most discussed approaches to the text, but are also concerned with very controversial and influential issues. By examining these two themes in particular, the present research aims to contribute to a current and important discussion in Conrad criticism. Second, "Africa and its representation" and "race and racism" lend themselves to a discussion of linguistic features. Both of them can be investigated through the analysis of the textual representations of two related aspects of the fictional world, i.e. Africa and Africans. Moreover, as I have argued in Chapter 1 (Section 7), the linguistic analysis of textual representations can also shed light on ideologies and stances beyond lexical choices. Finally, the above discussion has shown that these themes are debated and disputed, especially the issue of racism. There is no unanimous critical agreement about whether HoD is a racist text or not. Within the aim of the present research, such conflicting stances can be likely signals of discrepancies among the TTs, as different positions on controversial matters can result in different ways to deal with given features in the ST. In other words, the translators may have approached the translation differently according to whether they agree with Achebe (1990) or not, and whether they read HoD as imbued with racist implications or not.

CHAPTER THREE

Texts, corpora, methods: A corpus approach to literary translation

1. Introduction

This chapter will introduce the texts, the corpora, and the methods employed in the analysis. From a more theoretical point of view, I will discuss how corpus stylistics can be applied to the study of literary translation in order to develop a methodological framework that could be further employed in future analyses. I will start with the research questions in Section 2, while, in Section 3, I will present the ST and the TTs under investigation and introduce the two reference corpora used in this study. Section 4 is the core of the chapter, in which I will discuss the analytical steps that will be employed to answer the research questions. The discussion is both methodological and theoretical as it not only concerns the actual procedures to be used in the analysis chapters, but it also addresses the implications of applying such methods to the study of translation. Finally, in Section 5, I will close the chapter with some concluding remarks.

2. The research questions

The overarching aim of this thesis is to explore the role of linguistic patterns as building blocks of the fictional world and triggers of major themes, both in literature and in translation. It seeks to discuss the effects that textual patterns can have on the interpretation of the text and how this interpretation can change as a consequence of translation. This aim is realised by means of the corpus-assisted study of HoD and four of its Italian translations, examining how the lexical level of the texts contributes to literary meaning through repeated lexico-semantic patterning. Specifically, this study will answer the following four research questions, into which the overall contribution of this research has been broken down:

- 1) Which sets of text-specific words reflect the themes “Africa and its representation” and “race and racism” in HoD?
- 2) What are the linguistic patterns of these words and which functions do they develop?
- 3) What are the effects of translation on the linguistic patterns and their functions, and on the way they convey themes in the Italian versions?
- 4) How do the TTs relate to each other when examined as whole texts, as opposed to the way they relate to each other when only the identified linguistic patterns are compared?

The first is a key question in corpus stylistic research because it deals with the fundamental link between linguistic scrutiny and literary appreciation (see Chapter 1). Corpus stylistics approaches the analysis of a literary text through the examination of its linguistic form. Therefore, it is of primary importance to be able to link the linguistic level to the literary interpretation of the text, so as to ensure that the given theme, topic, or interpretational aspect of the text one wants to focus on is approachable from a linguistic point of view. The answer to the first research question will show how the two major themes selected are linguistically instantiated, in this way allowing me to identify formal features to study further as reflections of these literary themes.

The second question builds on the first one to examine in detail the patterning of the formal features that contribute to conveying the major themes. Conrad is renowned for his characteristic narrative voice (Senn 1980), the result of a constant stylistic effort that would enable him to “make [his readers] hear, to make [them] feel – it is, before all, to make [them] *see*” (Conrad 1950: ix-x, emphasis in the original). Once the relation between the two major themes and the lexical level of HoD has been established, by answering this research question I will identify the exact linguistic techniques Conrad uses to “make us feel, to make us see” his fictional world.

Specifically, this question explores the stylistic features that contribute to constructing the fictional representation of Africa and the fictional representation of Africans.

With the third research question, the Italian translations are brought into play. This question is concerned with the effects of translation on the interpretation and reception of literature. Translation can in fact reflect the translators' personal interpretation and/or bias, which in turn can affect the TT's reception in the target culture. By answering this question, I will explore the effects of the translational shift on the stylistic features previously identified. The fictional representations of Africa and of the African natives will be compared in the TTs and with the ST. The comparison aims to show how alterations at the textual (linguistic) level can affect the overall extra-textual (interpretational) level of the TTs, sometimes with manipulative effects on the text reception.

The purpose of the fourth question is to add a wider perspective to the comparison of the translations. The analysis of the TTs, as addressed by the third research question, focuses on comparing the effects of translation on the specific linguistic patterns identified through the second research question. In contrast, the fourth question aims to look at the TTs as whole texts, so as to examine whether their overall degree of similarity and difference differs from what emerged when the comparison was based on the specific linguistic features only.

In addition, by addressing these four research questions, the present study aims to make two complementary contributions. The first is offering an original contribution to the study of HoD, providing new perspectives on the short novel that cannot be obtained without a corpus-assisted method. The second one is building up a sound methodology for the study of literature in translation in general. This research seeks to develop a method that can be extracted from the specific study of HoD and its Italian translations and can be applied successfully to other works of literature and their translations.

3. Texts and corpora

The raw material for any corpus stylistic research is digital texts. These texts can either be studied individually, or gathered together to form a 'corpus', i.e. a collection

of digitalised texts selected with explicit criteria and for linguistic purposes so to be representative of a specific aspect of the language (for other definitions see Stubbs 2002, Tognini-Bonelli 2001, Fischer-Starcke 2010). The present study makes use of both individual texts and corpora, the former as object of the analysis, the latter as representative norms against which the object of the analysis is compared. Specifically, this research uses five individual texts, HoD and its four Italian translations, and two reference corpora, the general Reference corpus and the Conrad Corpus. The following sections will introduce and describe them, illustrating the criteria with which they have been selected and compiled, and explicating the aims for which they will be employed.

3.1 *Heart of Darkness*

Selecting the object of the analysis is never a straightforward choice, especially in corpus studies, where the selection is influenced by many external factors and at the same time has numerous methodological implications. Here as well, the selection of HoD is not fortuitous, but rather motivated by various methodological and practical reasons.

As I mentioned in the introduction, there are many reasons behind the decision to concentrate on HoD. The first is its popularity; Conrad is one of the most well-known authors in British and world literature and has been considered as part of the literary canon for decades (see Chapter 2). As a result, he has been extensively studied. The critical and academic interest in his works has been continuous since he started writing and HoD specially has taken centre stage in an innumerable amount of studies. This century-long interest is beneficial in many respects. First, because it provides plenty of material with which to inform a corpus stylistic analysis. This study investigates the linguistic instantiation of major themes in HoD. I have argued in Chapter 2 that the review of the literature can offer an informed picture of the themes mostly discussed by literary critics, emphasising those aspects of the texts that are central in the critical debate. Having a wide range of critical studies from which to take suggestions is therefore an important advantage for the present work. Second, the abundance of critical material spread over many decades is fundamental to

contextualise the text's reception at different stages. This diachronic perspective is important in order to frame the TTs in the critical milieu in which they have been produced. Translations respond to the socio-cultural environment in which they are born (Chapter 1, Section 7). Having a wide range of critical readings for each time period facilitates the search for the potential influences of the context of production on the TTs. Finally, Conrad's popularity is also mirrored in the large number of languages his texts have been translated into. HoD is the first work by Conrad that arrived in Italy in book format and since then it has become Conrad's most translated work in Italy (Curreli 2009: 149), with 20 different versions spanning from the first 1924 book translation to the latest editions of the 2000s. This allows me to select translations of different periods and analyse the potential effects that the ever-changing reception of the original has had on them.

The other main reason to choose HoD is purely practical, namely the availability of texts in digital format. One of the greatest limitations of corpus studies is that the range of potential analyses is influenced by the actual availability of digitalised texts (Fischer-Starcke 2010: 30). The programs for linguistic analysis work only with digitalised texts, so their availability is a critical factor to take into account when choosing what to study. Of course, texts that are not already available can be digitalised for the purpose. However, this process is often very demanding and time-consuming (see Section 3.3 below), or simply not possible at all due to copyright constraints. Hence it is not surprising that the choice of what to study is based also on what is already available in electronic format, with all the limitations in terms of language, genre, and historical period varieties that this availability implies. These limitations have been acknowledged by Fischer-Starcke (2010: 29), who admits that the original idea of including in her Gothic reference corpus all of the texts mentioned in Austen's *Northanger Abbey* was abandoned because these texts are not all available in electronic form. Similarly, Mahlberg (2013: 3) explains that the free availability of electronic texts is one of the three reasons for her choice to concentrate on Dickens.

The vast majority of Conrad's works, including HoD, is available for free from the website Project Gutenberg.³ As far as the translations are concerned, the high number of Italian versions is equally mirrored in the wide availability of many different e-book editions. These e-books provide a sufficient assortment of

³ <https://www.gutenberg.org/> (last accessed November 2014).

translations that do not need to be converted, being already in digital form. This means that all the texts needed for this research are already available in electronic format, except for the 1924 TT, which I have manually digitalised (see Section 3.3). The digital version of HoD used in the present study has been retrieved from Project Gutenberg. I have edited the file manually to remove the index, introduction, chapter headings, Project Gutenberg licence notes, and any other element that was not written by the author and/or makes no part of the body of the short novel. The resulting text counts 38,759 words.

3.2 The reference corpora

The term ‘reference corpus’ here refers to a corpus designed and compiled to be representative of a specific language variety so that it can be used as the basis for comparison. This study employs two reference corpora to be compared with HoD: the general Reference Corpus (RC) and the Conrad Corpus (CC). The main reason for employing them is to identify words that are typical of HoD, i.e. words with a deviant frequency. This will be achieved through a keyword analysis (see Section 4.1 for more details) and both reference corpora will be used to obtain two different keyword lists (Chapter 4). In addition, RC and CC will be also used in Chapters 5 and 6 to compare the frequency and usage of other relevant items, whether they are keywords or not, in order to provide an idea of how these items are specifically used in HoD as opposed to the other novels by Conrad (CC) or other fictional texts of the same period (RC).

RC aims to represent the English fictional written language contemporary to Conrad. Comparing HoD against RC will provide information about the differences between HoD-specific language and the more general literary language used in that period. RC is composed of 42 texts by 21 British authors, each author represented by 2 works, for a total of 4,177,872 words (see Appendix 1 for the list of texts used). These novels, short novels and collections of short stories were written between 1887 and 1922, in this way matching the years in which Conrad wrote his works: Conrad published his first work, *Almayer's Folly*, in 1895 – although he started writing it some years before – and his last novel, *Suspance*, was published posthumously in

1925. All the texts included in RC have been retrieved from Project Gutenberg. It is worth mentioning that RC has no internal balance between the texts. Although each author equally contributes to the corpus with two works, these have different lengths and therefore influence the data differently. This issue is dealt with by compiling the corpus using separate text files for each work, instead of a single one containing all the works together. In this way it is possible to spot whether one text – or a minority of them – is individually affecting the outcome of the analysis.

CC collects most of Conrad's fictional works, so that comparing HoD against CC will show the differences between HoD-specific language and Conrad's general fictional language. CC counts 1,724,568 words and includes all the fictional works by Conrad available on the Project Gutenberg website except HoD, for a total of 23 works: 18 novels and short novels and 5 collections of short stories (see Appendix 1 for the list of texts used). Conrad's collaborations with Ford Madox Ford and his autobiographical writings have been excluded, for this corpus is designed to study Conrad's fictional style, rather than his general written language.

All the texts in the two reference corpora have been manually checked to delete what does not constitute Conrad's fictional language: indexes, introductions, prefaces, notes, chapter headings, appendices, Project Gutenberg licence notes, and any other writing extraneous to the very body of his work.

3.3 The translations

This study makes use of four different Italian translations. Three of them are contemporary translations, whereas the fourth is an early translation, dated 1924. The texts of the three contemporary translations have been obtained from their e-book versions, available on various online e-book stores. The files have been converted into txt format and edited so as to delete their introductions, prefaces, indexes, notes, chapter headings, etc., in line with the editing of the ST and of the reference corpora. The Bur edition of HoD, *Cuore di Tenebra* (henceforth Translation B), was published in e-book format in 2010, but the translation was originally written by Giorgio Spina in 1989. It is composed of 37,715 words. The Garzanti edition of HoD, *Cuore di Tenebra* (henceforth Translation G), was published in e-book format in 2011 and

originally translated by Luisa Saraval in 1990; this text is composed of 38,493 words. The Mondadori edition of HoD, *Cuore di Tenebra* (henceforth Translation M), counts 36,442 words; it was published in e-book format in 2010, while it was originally translated by Rossella Bernascone in 1990.

The early TT (henceforth Translation S) is the 1924 translation by Alberto Carlo Rossi, the first Italian book-translation of HoD. The edition used for the analysis is the one published in 1928 (40,556 words) by the publisher Sonzogno, which reprinted the original 1924 translation with only some minor, formal alterations (Curreli 2009: 189). Translation S is not available in digital form so, after receiving the permission from the current copyright holder, I created a digitalised version of the text. To begin with, the book was scanned. Then, with the aid of *ABBYY FineReader*, an optical character recognition (OCR) software, the scanned pages were converted into a text file. This text went through a twofold process of checking. First, it has been automatically checked with the help of *Microsoft Word* spell checker; second, I have double-checked the text manually, in order to eliminate the remaining typos and misspellings. The OCR software, for example, is likely to confuse *l* with *I*, or *c* with *e*, especially when the scan of a page is not particularly clear. When the mistake of the OCR results in a non-existing word, *Microsoft Word* spell checker is able to spot it, facilitating the editing process. However, when the reading of the OCR does not produce what the spell checker sees as a mistake, for instance with *case/ease*, manual checking is required. This example gives an idea of the laboriousness of manually converting texts in electronic form: it is a long and complex process that involves repeated readings of the text, all of which require time-consuming precision. Table 3.1 provides details about the reference corpora and texts used in the present thesis.

Table 3.1 Texts and corpora

Text	Translator	Publisher	Translated	Published	Words
HoD	/	Blackwood	/	1902	38,759
Translation S	A. C. Rossi	Sonzogno	1924	1928	40,566
Translation B	G. Spina	Bur	1989	2010	37,715
Translation G	L. Saraval	Garzanti	1990	2011	38,493
Translation M	R. Bernascone	Mondadori	1990	2010	36,442
RC	/	/	/	/	4,177,872
CC	/	/	/	/	1,724,568

4. Data analysis

Corpus analysis results from the interaction of four fundamental components: ‘corpus creation’, ‘discovery’, ‘hypothesis formation’, and ‘testing and evaluation’ (Laviosa 2002: 8). These components interact with and inform each other in order to assemble the methodology of the study. In this process, the data that the corpus provides plays a cardinal role. However, how the data is used in relation to the hypothesis formation can vary. Borrowing Laviosa’s (2002) terms, the hypothesis formation does not necessarily have to be the successive step to the discovery (corpus creation > discovery > hypothesis formation > testing and evaluation), but can also precede the entire process (hypothesis formation > corpus creation > discovery > testing and evaluation). In the first case, the data resulting from the discovery is the starting point for the hypothesis formation, which is then tested and evaluated. In the second case, the hypothesis is formulated before the corpus creation and the discovery, and thus informs them. The data resulting from the discovery is used to test and evaluate the pre-existing hypothesis. These two different approaches to corpus analysis have been called respectively ‘corpus-driven’ approach and ‘corpus-based’ approach (Tognini-Bonelli 2001). Tognini-Bonelli (2001) defines ‘corpus-based’ as the approach that interrogates the corpus to testify and improve pre-existing hypotheses “that have not themselves been formulated in the light of corpus data” (Tognini-Bonelli 2001: 65). Whilst a ‘corpus-driven’ approach (Tognini-Bonelli 2001: 84) aims at building hypotheses from scratch, completely free from theoretical premises, so to reflect entirely the corpus evidence.

As far as this research is concerned, the distinction between corpus-based and corpus-driven approach has a mainly heuristic value, as in practice the two approaches can be combined. This has been practically proved by Rayson (2008), for example, who combines elements of the corpus-based and the corpus-driven paradigms into his ‘data-driven’ approach (Rayson 2008: 543). The combination is possible because the opposition between the two approaches is more theoretical than realistic. In fact, the main characteristic that distinguishes the corpus-driven from the corpus-based approach, i.e. the theory-free paradigm, is “a myth at best” (Gries 2010: 330), as achieving a complete detachment from any existing assumptions is virtually impossible. More realistically, the corpus-driven approach aims to keep theoretical

premises at the minimum, so the difference between the two paradigms can be seen in terms of the degree to which they employ theoretical premises, rather than whether they use them or not.

In line with this, the present study does not comply with any specific approach in particular, but rather adopts a more flexible attitude (hence the choice of “corpus-assisted” in the title of this thesis). Corpus-based is of course the dominating approach in this analysis, given that the review of the literary criticism in Section 2 has provided me with indications and suggestions for the investigation of the data. However, this does not prevent the use of other methods or approaches that tend towards the -driven end of the corpus-based/corpus-driven continuum. Specifically, in Chapter 7 I will use principal component analysis to compare the TTs as whole texts, as opposed to the comparison based on selected stylistic patterns undertaken in Chapters 5 and 6. As such, the analysis in Chapter 7 will be more corpus-driven than corpus-based, since the criteria for the comparison will not be based on pre-existing premises, e.g. derived from the literary criticism. On the contrary, the texts will be compared purely on the basis of the frequency patterns of their most frequent words, so the findings will reflect only the formal features of the texts, aiming to keep pre-existing hypotheses at the minimum.

The methodology of the present research is structured in four steps, as represented in Figure 3.1. I will briefly introduce them here in order to have a complete overview of the methodology, but I will focus on each of them separately in the remaining sections of this chapter. The first step (Chapter 4) identifies groups of text-specific words that are related to the major themes discussed in Chapter 2. This step moves the focus of the analysis from the whole text to specific items, in order to narrow down the range of the investigation to individual items that are relevant to the purpose of the present work. To do so, I will carry out a keyword (KW) analysis, with the aim of emphasising the link between the lexical level of HoD and its major themes. The selection of specific groups of KWs to analyse in detail will be made adopting Mahlberg & McIntyre’s (2011) model for keyword categorisation. This particular model is chosen because it is specifically designed to be applied to literary texts and allows the researcher to focus on the recognition of themes rather than on the counting of KWs. This first step of the methodology will provide an answer to Research Question 1.

The second step of the methodology examines in detail the lexico-semantic patterns of the words in the categories selected, in order to investigate their functions as building blocks of the fictional world. In particular, Chapter 5 is concerned with the functions of patterns that convey the fictional representation of the African jungle, whilst Chapter 6 deals with the functions of patterns related to the fictional representation of the African natives. Quantitative and qualitative analyses are linked in these chapters, where the items are studied as units of meaning (Sinclair 1991, 2004) which develop specific local textual functions (Mahlberg 2007b, 2013). In this investigation, semantic preference and semantic prosody will take centre stage, as they play a fundamental role in constructing linguistically the jungle and the natives in HoD. This step will provide an answer to Research Question 2.

The third step of the methodology explores the effects of translation on the lexico-semantic patterns and their functions identified in the original. This step too is carried out in Chapter 5 and Chapter 6: the TTs will be compared to highlight the differences and similarities in the fictional representation of the African jungle and of the African natives. This comparison will show the consequences of textual alterations on the potential interpretation of HoD in the target context, with a special emphasis on the manipulative effects such alterations may convey. This step will provide an answer to Research Question 3.

In the last step of the methodology, principal component analysis is used to compare the TTs to each other and with the ST. Here the perspective on the texts' relation is shifted from the level of individual linguistic features to that of the whole texts. This step aims to counterbalance the previous comparison and to account for the TTs in their entirety. This procedure will provide an answer to Research Question 4.

Step 1	Identification of specific words related to major themes	Answer to RQ1
Step 2	Detailed analysis of word patterns and their functions	Answer to RQ2
Step 3	Comparison of patterns and functions in translation	Answer to RQ3
Step 4	Comparison of the translations as whole texts	Answer to RQ4

Figure 3.1 Methodology overview

4.1 Lexis and themes

By its very nature, the corpus approach prioritises the lexical level over the other levels of a text (Mahlberg 2005: 189, cf. also Chapter 1). This may be seen as a limitation, especially in the study of literary texts, as lexis is only one of the levels through which the ‘literariness’ (Carter 2004) of a text is conveyed. To compensate for this, it is particularly important in corpus stylistics to relate patterns of lexis to their functions, so that the analysis of the lexical level is not limited to the linguistic description, but also encompasses the literary appreciation of the text (cf. Chapter 1). In other words, the corpus stylistician should use the insights the corpus provides as a means towards a deeper understanding of the literary work. This emphasises once more the importance of the link between linguistic analysis and literary interpretation, and one way to establish this link is by focusing on lexical choices that are stylistically relevant.

The search for relevance in lexical choices is benefitted by the fact that lexis is the level of linguistic form which allows for the greatest variability and the greatest freedom of choice (Fowler 1966: 16). Thus, the selection of one word instead of another is likely to convey a certain degree of meaningfulness. However, not every lexical choice forms part of the style of a text indiscriminately. Generally, to be considered a feature of style, a lexical choice has to be consistent: choices in isolation may be stylistically meaningful *per se*, but it is patterns of repeated choices that develop the style of a text as a whole (Leech & Short 2007: 34). Consistency and repetition in lexical choices are best examined in terms of their frequency in the text, so frequency can be a useful indicator of stylistic relevance (cf. Chapter 1). Adopting Leech & Short’s (2007) notion of style as a function of frequency, any frequency deviance has the potential of conveying literary relevance, given that a linguistic feature must be deviant in order to be foregrounding (Leech & Short 2007: 41). Therefore, the first step towards the search of foregrounding features is comparing the actual frequency of a given lexical choice against its expected frequency, in order to establish whether the frequency of this lexical choice is deviant or not. This is not to say that all deviances are stylistically significant, but rather that deviant features are likely candidates to convey literary relevance, as opposed to features with a normal frequency of occurrence.

In corpus linguistics, a common way to identify words whose frequency is unusually high (or low) in comparison with some norm is through KW analysis. Within the aims of corpus stylistics, the use of KW analysis has a twofold value: it enables an easy and automatic identification of deviant words and, at the same time, it establishes a link between lexis and the content of the text. In fact, KWs provide an effective way to characterise a text or a genre (Scott 2015a: online). Scott (2015a: online) explains that “aboutness” KWs – items that indicate what a text is about – are one of the three typical categories of KWs. Similarly, Stubbs (2005: 11) recognises that nouns from a KW list can reflect the topics of a text.

However, Stubbs (2005: 11) argues that only superficial topics are emphasised with KW analysis, while “underlying themes” are not visible. In the case of HoD, for example, he explains that noun KWs such as *Kurtz* and *river* indicate only superficial topics, as opposed to inner themes. I argue instead that KWs can be related to underlying themes as well as to superficial topics. Although individual KWs alone cannot account for the totality of a theme in a literary text, I will show that groups of KWs in HoD are linked to its major themes. This link is established through the patterns that the KWs create. The relation between KWs and fictional themes is also acknowledged by Mahlberg & McIntyre (2011), who maintain that KWs have the potential to be both “signals for the building of fictional worlds” and “triggers for thematic concerns of the novel” (Mahlberg & McIntyre 2011: 207). Toolan (2008: 112) as well raises the same point, explaining that KWs can be “pointer to a text’s themes and preoccupations”. KW analysis can therefore help the corpus stylistician to relate linguistic description and literary interpretation. Given this potential to relate deviant items and content, KW analysis is adopted in the present study to address the first research question: how to identify sets of text-specific words that reflect the themes “Africa and its representation” and “race and racism”.

From a quantitative point of view, carrying out a KW analysis is beneficial in many respects. First, it is an efficient way to begin a study, because it substantially narrows down the data to analyse and facilitates the selection of a fewer items to examine in detail. At the same time, the procedure circumvents the researcher’s bias, as KWs are generated automatically by the computer, “blindly”, without considering meaning relations between them. Moreover, all KWs are statistically significant by definition, as they are the result of a calculation that reduces the risk that their deviant

frequency is due to chance. In sum, KW analysis allows the researcher to start the investigation with a controlled amount of items that are statistically significant and automatically generated.

These advantages have been shown, for example, by Culpeper (2002b: 14). He explains that KW analysis is particularly beneficial because it is an automated technique that, within minutes, emphasises words that are good indicators of potentially relevant features of a text. He demonstrates this by looking at patterns in the characters' dialogues in Shakespeare's *Romeo and Juliet*, identifying both content and function words that characterise the style of the protagonists' talk. Specifically, the KW analysis emphasises, on the one hand, KWs such as *beauty, blessed, love, dear, stars, and fair* which are explicit signals of Romeo's "love talk", matching one's intuition about the character (Culpeper 2002b: 20). On the other hand, the analysis also emphasises KWs such as *if* and *yet* that are not immediately recognisable as characteristic of Juliet, yet they indicate a relevant feature of her style, that is, the anxieties of Juliet's talk.

Toolan (2008) uses KWs in his analysis of narrative prospection and expectations. He assumes that some textual features create expectations for the progression of the narrative, and that these features must be prominent in order to be identified by the reader and create an expectation (Toolan 2008: 108). He therefore uses KW analysis to specify foregrounded features of texts. He explains that KWs are particularly useful for his study, because they are both thematically relevant and unlikely to be ignored by the reader, by virtue of their recurrent frequency. Čermáková & Fárová (2010) prove that KW analysis can be similarly useful in the study of literary translation. They look at KWs in J. K. Rowling's *Harry Potter and the Philosopher's Stone* and its Finnish and Czech translations. They aim to investigate how some thematically relevant KWs (*wand, cloak, owl, eyes, door, wizard, broomstick, troll, broom, and stone*) have been translated. Čermáková & Fárová (2010: 183), through qualitative analysis of text extracts, compare the functions of the KWs in the ST and in the TTs, concluding that a KW analysis could have helped the translators to identify more easily items that required special attention.

It is worth specifying that generating a KW list does not constitute an analysis in itself. Despite the advantages of using this technique, it is the qualitative

examination of KWs that points out their meaningfulness. This is particularly true in the context of corpus stylistics and the study of literary texts, where KWs have to be examined in light of the text content and its critical interpretation. O'Halloran (2007), for instance, focuses on the KW *would* in James Joyce's "Eveline". This item was recognised as a KW in the same text by Stubbs (2002), but Stubbs's (2002) qualitative analysis of the function and patterns of *would* is very limited. Stubbs (2002: 130) simply states that *would* signals the fact that in the short story Eveline is thinking about hypothetical possibilities for her future, although in the end she fails to put them in practice. In contrast, O'Halloran (2007) investigates the way *would* is used in the short story both quantitatively and qualitatively, exploring its patterns in free indirect thoughts (for example *she would [...]* and *Frank would [...]*) using Hallidayan transitivity analysis (Halliday & Matthiessen 2004). O'Halloran (2007: 236) to prove that Eveline's conscious hopes about her future leak instead subconscious doubts and signal in advance her final decision not to follow Frank.

The main purpose for adopting KW analysis in the present research is to identify HoD-specific words that relate to the major themes discussed in Chapter 2. As I aim to relate the lexical level of HoD to its themes, in the present study KWs will be seen mainly as building blocks of the fictional world and triggers of thematic concerns. To address this specific purpose, my study adds a further level of discrimination to the KW analysis, a classification into categories that are likely to represent the themes of HoD better than individual KWs. The classification of KWs into sets and categories is not new in corpus linguistics. McEnery's (2009) develops a model to identify moral panic in a corpus through the classification of its KWs. He identifies seven categories that would reflect a moral panic in a text or a corpus: "object of offence", "scapegoat", "moral entrepreneur", "consequence", "corrective action", "desired outcome", and "moral panic rhetoric". If the KWs resulting from the comparison of two corpora fit this categorisation model, than this indicates that the corpus under analysis contains a moral panic (McEnery 2009: 96). He puts his model into practice analysing the writings of Mary Whitehouse between the 1960s and 1970s, exploring her ideological stances on a range of different topics such as school system, pornography, and decency in general.

Similarly, Gabrielatos & Baker (2008) adopt KW analysis to explore discursive constructions around refugees, asylum seekers, immigrants, and migrants

(RASIM) in the UK press between 1996 and 2005. They argue that by grouping relevant KWs into specific topics and attitudes it is possible to identify the different discourses of RASIM between tabloids and broadsheets (Gabrielatos & Baker 2008: 10). Comparing the categories, they show that, for instance, tabloids focus on the number of RASIM, on the manner and place of entry, and on the alleged abuse they commit. Broadsheets instead give the topic a more in-depth treatment, discussing international, social, political, and religious issues (Gabrielatos & Baker 2008: 30). A similar application is Bachmann's (2011) analysis of discourses on civil partnership in the UK Houses of Parliament. Bachmann (2011: 88) as well divides KWs into thematic groups and then investigates their functions in context, demonstrating how this methodology allows him to emphasise which topics are predominant in the debates (Bachmann 2011: 100).

However, these models have been used to study language varieties other than literary language, and therefore may not work equally well for the analysis of HoD. The model here adopted is instead specifically designed to work with literary texts, namely Mahlberg & McIntyre's (2011) model. Mahlberg & McIntyre's (2011) model offers a way of classifying relevant KWs into thematic categories which emphasise their role as building blocks of fictional worlds and themes. They propose two main categories: 'fictional world signal' KWs and 'thematic signal' KWs. In Chapter 4, I will show how these sets fit perfectly the aims of this research and help to establish a link between the lexical level of HoD and its major themes.

4.2 From words to extended units of meaning

Firth (1957: 179) famously stated that "you shall know a word by the company it keeps". Words in fact do not exist in a void. On the contrary, they establish meaningful connections with the other words that precede and follow them. Knowing these connections is fundamental for the stylistician, because the difference between an unusually frequent feature and a feature with stylistic relevance may depend on the connections such a feature establishes with its surrounding context. In other words, a deviant feature can be seen as contributing to a style and possessing literary relevance only when it "form[s] a significant relationship with other features of style, in an

artistically coherent pattern of choice” (Leech & Short 2007: 40). The relationship between words is therefore fundamental for the creation of artistically motivated style: the study of stylistic features should go beyond the analysis of individual words and encompass the relationships that words establish with each other.

More than anyone else, Sinclair (1991, 2004) has shown that focusing only on individual words is limiting, because the word in itself is not an independent unit. A more suitable unit of meaning is the ‘lexical item’ (Sinclair 1991, 2004), a functionally complete unit that overcomes the formal boundaries of the single word. The lexical item is defined by the sum of the lexical, grammatical, semantic, and pragmatic relations the item establishes in its context of use. Sinclair (1991, 2004) proposes a model to study the lexical item as a unit of meaning, a model that revolves around the interaction between the ‘core’ item and four structural categories: ‘collocation’, ‘colligation’, ‘semantic preference’, and ‘semantic prosody’. Meaning is likely to develop from the interaction between the structural categories and the core.

The first structural category develops on the lexical level and it is called ‘collocation’. Collocation (Sinclair 2004: 28) is defined as the tendency of two or more words to occur frequently together within a few words (usually four or five) of each other. It is a purely lexical relation based on probabilistic principles (Stubbs 2002: 64) which, although having a limited effect on the overall meaning of the unit, operates strong constraints at the level of lexical choice (Sinclair 2004: 28). This means that the selection of a word can be limited by its collocational patterns. Moving to a higher level than the lexical one, we find colligational relations. ‘Colligation’ (Sinclair 2004: 142) develops on the grammatical level and is defined as the tendency of grammatical phenomena (or lexis plus grammatical phenomena) to co-occur together. Colligation is a more abstract category than collocation, for its identification involves conceptualising words (immediately perceptible) as representing grammatical categories.

The next step of the model moves from individual words and grammatical categories to groups of semantically related words. This semantic interaction between the core and its context is termed ‘semantic preference’. Sinclair (2004: 142) defines it as the tendency of an item to co-occur regularly with semantically related words, i.e. words forming part of the same semantic field. Semantic preference too moves forward the model towards abstraction (Stubbs 2002: 88), as semantic fields have

typical and frequent members, but they mainly remain open-ended categories. The last structural category, ‘semantic prosody’, extends to the pragmatic relations of the unit of meaning and contributes to express the speaker/writer’s attitude. It is defined by Sinclair (2004) as the attitudinal tendency in the realm of pragmatics that unveils evaluative stances about the topic of the discourse. It is the most abstract structural category, because it is open to a wide variety of different realisations, since “in pragmatic expressions the normal semantic values of the words are not necessarily relevant” (Sinclair 2004: 34). Nevertheless, semantic prosody plays a key role in the construction of the unit of meaning as it defines the function the unit is meant to have. All the other categories of the unit follow this functional choice and are interpreted accordingly.

Let us consider a practical example. Figure 3.2 below shows 30 concordance lines of *reduce* sorted randomly, retrieved via the online tool *BNCweb*.⁴ Concordance lines enable a “vertical” reading of the corpus which facilitates the identification of repeated patterns (cf. Tognini-Bonelli 2011: 3). At the collocational level, *reduce* co-occurs frequently with *costs* (*reduce the costs*, *reduce its costs*, *reduce costs*, *reduce real payroll costs*, and *reduce costs further*). *Costs* is in fact one of the strongest collocates of *reduce* in the BNC, together with *emission* and *risk*, in a span of 5:5 (5 words at the left and 5 words at the right of the core *reduce*). *Costs* is not the only noun occurring at the right of *reduce*, however. This core co-occurs frequently with other nouns or noun phrases, for example *reduce the myth*, *reduce agricultural subsidies*, *reduce evaporation*, *reduce the total fat*, *reduce the deep-felt sense of dissatisfaction*, *reduce the heat*, etc. This tendency is an instance of colligation: *reduce* colligates frequently with nouns and noun phrases at the R1 position (one slot at the right of the core).

Grouping together these nouns and noun phrases, it is possible to identify the semantic preference of the core *reduce* for words belonging to the semantic fields of business/economics (*costs*, *strategy*, *taxpayer*, *solicitors*, *debt*, *payroll*, *company*, etc.), medicine (*fat*, *diet*, *urine*, *surgical*, *operations*, *abdominal*, *cholesterol*, etc.), and environment/pollution (*evaporation*, *petrol*, *consumption*, *pollution*, *emissions*, etc.). From a pragmatic point of view, it may seem that *reduce* have a negative meaning. However, the concordance lines show that it has instead an overall positive

⁴ <http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php> (last accessed June 2015).

semantic prosody. This is mainly due to the fact that the objects of the verb *reduce* are things which are usually considered beneficial to reduce, such as *reduce the cost*, *reduce significantly its overall operational impact on the environment*, *reduce its outstanding debt*, *reduce the total fat*, *reduce petrol consumption*, *reduce conflict*, *reduce weight*, *reduce your cholesterol levels*, *reduce pollution emissions*.

As mentioned before, the pragmatic function of the lexical item dictates the other functional choices of the unit, as the semantic prosody and the core are the obligatory elements in Sinclair’s (1991, 2004) model. In this case, the positive semantic prosody of *reduce* calls for collocates that can convey this positive evaluation. Hence *reduce* often refers to things or conditions that it is good to decrease because, for example, they are dangerous for one’s health, business, or the environment: among them, the most frequent are *costs*, *risk*, and *emissions*. Grammatically speaking, these things and conditions are expressed with the use of nouns or noun phrases, hence the colligational tendency of *reduce* to co-occur with this particular grammatical class.

ayer by raising the revenue from users; deterrence — to	reduce	the cost to the taxpayer by reducing demand for a service;
the white man has felt compelled to tame the desert, to	reduce	its primeval vastness to human dimensions. In particular, I
ion of abstract and general relations. These two features	reduce	the myth to the level of ‘theory’ as far as
ommunication networks within the processor complex to	reduce	overhead in communication between the processors. Peak
Agriculture, Clayton Yeutter, urged the EC and Japan to	reduce	agricultural subsidies. At the end of October the Departm
ssues. The Bank has developed a strategy which aims to	reduce	significantly its overall operational impact on the environ
system to sustain, and that is why he took steps to	reduce	it. That has been welcomed everywhere except on the Op
aber of the European Community. Fragmentation would	reduce	our influence to virtual nonexistence. And I also believe th
through the ages been obsessed with finding methods to	reduce	complex, individual human personalities to ‘types’? In me
tice permanently.’ Solicitors also fear the changes could	reduce	public access to legal aid and increase the likelihood of a
ver the next twelve months. DG is taking Luna boxes to	reduce	its costs — Omron is still seeking other OEMs. Omron:
t result in lower costs. Learning effects may also help to	reduce	costs and thereby prices. These effects arise when expansi
ld be a form of national savings. The government would	reduce	its outstanding debt, freeing capital for private investment
in there were behavioural remedies proposed in order to	reduce	the extent of the product tie. Following the report, althoug
will be obliged to water crops at dusk or dawn only to	reduce	evaporation. Watering gardens, washing cars and filling s
ut the total calories down to 1,850. The simple changes	reduce	the total fat in the diet by quiet a large amount.
money that you pay to the government and, second, to	reduce	real payroll costs while raising employees’ perceived inco
wanting to pass urine occurs. This approach attempts to	reduce	the frequency of passing urine and increase bladder capac
d with overdrive (product feature) which enables you to	reduce	petrol consumption on motorways’ (customer benefit). Th
simmental. However, breeders have been taking steps to	reduce	calving problems and also to breed longer, taller animals.
s which had taken place since 1979 had done nothing to	reduce	the deep-felt sense of dissatisfaction, and that in many est
ulsive overeaters or performing surgical Operations to	reduce	the abdominal apron or reduce the length of the intestines,
raise the supply price (S), which in turn will	reduce	the MEI. This could follow if the capital goods producing
arts? How is each item of the Country Code intended to	reduce	conflict between locals and visitors? How may the use of
. She was also advised that she should continue to try to	reduce	weight and should not return to her work (which involved
, and some 5000 are already fitted in homes. It can	reduce	the amount of water needed to flush a lavatory by 40 per
analysis. Even if one thinks one’s own company will not	reduce	costs further, it is dangerous to make the same assumption
cally. Healthier eating will help control your weight and	reduce	your cholesterol levels, so the simple message is: Eat less
n control options which can be implemented speedily to	reduce	pollution emissions from traffic, industry and homes. Dec
the eel (if using). Bring to the boil,	reduce	the heat and simmer for 10 minutes. 3 Add the remaining

Figure 3.2 Example of unit of meaning in the BNC: *reduce*

Sinclair's (1991, 2004) model has shown that meaning is a shared contextual endeavour, rather than an intrinsic feature of individual words. In a similar fashion, style too is a shared contextual endeavour, the sum of artistically coherent patterns of choices (Leech & Short 2007: 40), rather than the result of single linguistic features. Both meaning and style, then, emerge from the mutual interaction of linguistic features on different textual levels. This parallelism would make Sinclair's (1991, 2004) model suitable to analyse both the conveyance of meaning and the construction of style. However, in practice, this model has been designed to work with a large amount of data, and might not work equally well with short texts such as HoD. The lexical item model provides a detailed picture of how meaning is developed on mutually related linguistic levels (lexical, grammatical, semantic, and pragmatic), but I will show that literary effects in HoD can develop on fewer levels than all four of them, for example from the lexical or semantic level alone. In addition, Sinclair's (1991, 2004) unit of meaning is mainly based on a fixed core and the patterns around it, whilst I will show that in HoD stylistic effects can be shared and spread over more than one core, each contributing in a different but interrelated way to the same function. In other words, the structure of the lexical item and the way it works in theory may not match the way it practically works in a short literary text.

To tackle the potential mismatch, a more flexible notion is adopted in this study, a notion which has been designed to work also with literary texts and to cope with their specific traits: Mahlberg's (2007b, 2013) 'local textual function'. Local textual functions are 'textual' because they account for the function of items in texts, rather than in more generalised language contexts. They are 'local' because their description of functions is specific to the text(s) analysed, so do not necessarily have to work outside of it (Mahlberg 2007b: 121, 2013: 17). Most importantly, local textual functions can be shared or similar across multiple units, accounting for the interaction of different lexical items towards the creation of the same meaning/function. This addresses the case of different lexical items contributing individually to the unitary construction of the same theme or aspect of the fictional world.

The combination of Sinclair's (1991, 2004) lexical item model and Mahlberg's (2007b, 2013) local textual function constitutes the groundwork for the second step of the methodology. This step is applied in Chapters 5 and 6 to study in detail the textual patterns that contribute to the themes "Africa and its representation"

and “race and racism”. This combined application will provide an answer to Research Question 2. In particular, Chapters 5 and 6 will build on the findings of the KW categorisation in Chapter 4, and will explore how the fictional representation of the African jungle and the fictional representation of African natives are linguistically realised. The items forming part of the thematic categories will be studied as units of meaning, with particular attention paid to their mutual interaction in order to highlight shared and text-specific local textual functions. Chapter 5 and Chapter 6 will especially focus on two interacting phenomena: semantic preferences/prosodies and the cohesive network they establish throughout the text via lexical cohesion.

4.2.1 Semantic preference and semantic prosody in Heart of Darkness and its translations

The focus on lexico-semantic patterns is due to the central role semantic preference and semantic prosody play in the analysis in Chapter 5 and Chapter 6. The importance of semantic preference and prosody in corpus stylistics is due to the unity of effect they can establish across several items, overcoming the issues arising from working with smaller corpora and hence lower frequencies. The corpus stylistician often works with frequency ranges that are much lower than those which a corpus linguist is used to dealing with. HoD, for example, is in the range of a few dozen thousand words, as opposed to the multi-million (or even multi-billion) word corpora employed in corpus linguistic research. In principle, this is not a problem since frequency is always a relative parameter and its significance is based on comparison. However, in practice, low frequency ranges affect the potential of frequency to highlight repeated patterns in texts. Setting a frequency threshold and limiting the analysis to phenomena whose occurrence is higher than the cut-off point (e.g. collocational or colligational phenomena) can leave out part of the picture.

For example, Baroni & Bernardini (2003) and Bernardini (2007) advocate a frequency-based approach to the identification and study of collocations in translation. Their method aims to account for the strength of the relationship between the collocates, through the use of statistical calculations. Within the aims of the present study, this approach would be limiting. Consider a word like *darkness*, which occurs only 25 times in HoD. We might be aware of its relevance in the text, because it is a

KW and because it is intuitively recognisable as representing one of the novel's main themes; as such, we might choose it for further investigation. However, with only 25 occurrences, it is unlikely that examinations based only on frequently repeated behaviours highlight anything relevant. In fact, if we look at the collocations of *darkness*, only 7 collocates occur more than 3 times, and only 11 occur more than twice; among these 11 collocates, only 2 are content words. There would not be enough data to identify frequency-based collocational patterns that can be relevant to the aims of this study.

On the contrary, frequency cut-off points are no more a defining criterion for the analysis of semantic preference and semantic prosody, as both phenomena can be constructed by the cumulative contribution of words that co-occur with a core just once. In the case of *darkness*, all of the collocates that occur only once could still be relevant to the analysis. Although they would not contribute to collocational patterns, they would instead contribute to semantic fields. Moreover, semantic preference and prosody function over wider contexts, involving larger sets of words, and can be shared by different items: several units of meaning contributing together to the same semantic preference and/or prosody. This is in line with what Rayson (2008) shows in his study of the Labour and Liberal Democratic parties' manifestos for the UK 2001 general election. He proves that the identification of key semantic domains in short texts (the manifestos are around 30,000 words each) highlights aspects of the texts that previous frequency-based KW analyses ignored. Rayson (2008) argues that the investigation of dominant semantic fields has been useful to "group together lower frequency words and multiword expressions which would, by themselves, not be identified as key, and would otherwise be overlooked" (Rayson 2008: 544).

Although semantic preference and semantic prosody are phenomena which have been extensively discussed (in particular, Partington 2004 and Bednarek 2008), previous research has mainly focused on the English language, while other languages and contrastive analyses have been neglected (Xiao & McEnery 2006: 108). Among the few exceptions, Xiao & McEnery's (2006) study of collocation, semantic prosody, and near synonymy in English and Chinese shows the relevance of semantic preference and prosody in the context of translation. They focus on three sets of near synonyms in English: the *consequence* group, the CAUSE group, and the *price/cost* group. With quantitative data and qualitative analysis, they prove that even what is

considered the Chinese equivalents of the English terms can have different semantic prosodies. Being aware of this difference is helpful for the translator and the teacher of second language alike (Xiao & McEnery 2006: 126).

Similarly, Oster & van Lawick (2008) aim to raise awareness of the importance of semantic preferences and prosodies among translation students and professionals. Failing to reproduce semantic and pragmatic conventions can result in unintentionally imprecise translations. Oster & van Lawick (2008) discuss the example of the Spanish expression *tomarse algo a pecho* and the German equivalent *sich zu Herzen nehmen*, concluding that even though the two may have similar meaning, they are not pragmatically equivalent (Oster & van Lawick 2008: 342): the Spanish version is sometimes used in negative or offensive utterances, while the German equivalent is not. Munday (2011) also examines semantic prosody in translation in his contrastive analysis of *loom large* and its Spanish dictionary correspondent, *cernerse*. Munday (2011) discusses Louw's (1993) understanding of prosodies, according to which they are inaccessible to the speaker/writer's conscious intuition. Although inaccessible to the speaker/writer, Munday (2011) argues that prosodies could be perceived by the reader, who is able to grasp the irony resulting from contradicting an established prosody. Therefore, translators too may be aware of semantic prosodies, or at least the effect it creates in the text (Munday 2011: 182-183).

Building on the existing research, the present work will contribute to the discussion of semantic preference and semantic prosody in translation. With my analysis, I will look in detail at the effect of translation on these phenomena, comparing text extracts across several translations. I will also consider whether alterations to semantic preferences and prosodies can affect the building of the fictional world and the major themes in HoD. Overall, this research contributes to a more nuanced understanding of these aspects of the unit of meaning in the specific context of literary translation.

4.2.2 Lexical cohesion in Heart of Darkness and its translations

One of the main topics discussed in the following chapters (particularly in Chapter 5) is lexical cohesion. This focus is due to the fact that interacting lexical items will be

shown to create shared semantic preferences and prosodies across the whole texts. The lexical network these patterns establish contributes significantly to the cohesion of HoD. Given the importance of this aspect, before moving to the third step of the methodology, it is worth discussing further the concept of cohesion, how it has been approached with corpus methods, and how it has been discussed in the context of translation.

‘Cohesion’ is the sum of meaning relations existing within a text. It occurs when the interpretation of a given textual element is dependent on the interpretation of another one, so that decoding one effectively involves decoding the other accordingly (Halliday & Hasan 1976: 4). For example, in the previous sentence, *one* is interpreted in relation to *element*; as such, it establishes a cohesive link with it. This type of cohesive relation is defined ‘substitution’ and involves the replacement of one item by another (Halliday & Hasan 1976: 88). Substitution is one of the four types of ‘grammatical cohesion’, together with ‘reference’, ‘ellipsis’ and ‘conjunction’. In addition to the grammatical level, cohesion can also be realised at the lexical level, through the selection of vocabulary. In this case, we talk about ‘lexical cohesion’ (Halliday & Hasan 1976: 6).

In Halliday & Hasan’s (1976) model, there are two main lexical cohesive devices: ‘reiteration’ and ‘collocation’.⁵ Reiteration is created when one item refers back to another one with which it shares a common referent. It can be realised through the repetition of the same word or with the use of a synonym, near synonym, superordinate, or general word (Halliday & Hasan 1976: 278). Collocation is established by the association of words that regularly occur together and share any recognisable semantic relation (Halliday & Hasan 1976: 285). In contrast to reiteration, two items linked by collocation do not have to share the same referent. Collocation can in fact be created not only by synonyms, near synonyms, and superordinates, but also by complementaries (*boy/girl, stand up/sit down*), antonyms (*wet/dry, like/hate*), or converses (*order/obey*), as long as a meaning association is recognisable. Collocation is therefore an umbrella term that covers any instance of cohesion resulting from the co-occurrence of words that are associated by various types of meaning relations (Halliday & Hasan 1976: 287).

⁵ The use of the term ‘collocation’ by Halliday & Hassan (1976) differs from how Sinclair (1991, 2004) uses it. In this case, collocation does not refer to one of the four structural categories of the lexical item.

Lexical cohesion – and especially collocation – is always contextual, in the sense that it is the occurrence of a word in the context of other related lexical items that enables the word to function cohesively. This makes lexical cohesion more difficult to estimate than grammatical cohesion, because its effect on the text is subtle (Halliday & Hasan 1976: 288). Potentially, any lexical item can or cannot establish a cohesive relation, but there is no way to know it just looking at the item by itself. On the contrary, with grammatical cohesion it is relatively easier to say. Halliday & Hasan (1976: 288) explain that *he*, for example, clearly calls for a cohesive link, as its identity needs to be retrieved from somewhere. This does not necessarily happen with lexical cohesion, as a lexical item “by itself carries no indication whether it is functioning cohesively or not” (Halliday & Hasan 1976: 288). An important part of the study of lexical cohesion is therefore the analysis of the textual environment in which words occur, in order to investigate whether a cohesive link with other associated words in the context is established or not.

When Halliday & Hasan (1976) published their study, they claimed that the most important thing to do when analysing a text in respect of lexical cohesion is to use common sense and rely on our knowledge as speakers of the language and its vocabulary: “[w]e have a very clear idea of the relative frequency of words in our own language, and a ready insight [...] into what constitutes a significant pattern and what does not” (Halliday & Hassan 1976: 290). Today, besides our common sense and knowledge, we can also use corpus linguistics and its tools. Corpus methods provide more precise frequency figures and a better illustration of patterns in texts; hence they can contribute to an important degree to the study of lexical cohesion.

From a corpus linguistic perspective, lexical cohesion can be studied as the interactions of units of meaning and the cumulative effect these interactions create. For example, Cheng (2009) analyses lexical cohesion in a corpus of public speeches about the SARS outbreak in Hong Kong and shows that the semantic preferences and prosodies of related items interact to create a cohesive discourse. The patterns of seven interrelated items (*Hong Kong, health, care, hospital(s), private, SARS, and virus*) establish a complex network of lexical and semantic relationships that contributes to lexical cohesion, coherence, and finally meaning (Cheng 2009: 81). Mahlberg (2009) describes the interaction of units of meaning in terms of their local textual functions and looks at them as a way to study lexical cohesion. Local textual

functions play a fundamental role in connecting patterns of words, and this connection, in turn, creates continuity between parts of the text (Mahlberg 2009: 110). Finally, the corpus analysis of units of meaning and local textual functions can also add further detail to the study of lexical cohesion.

I have argued in Section 4.2 that units of meaning and their functions can be described in terms of the lexical patterns they create in the text. Hoey (1991) argues that “the study of the greater part of cohesion is the study of lexis, and the study of cohesion in text is to a considerable degree the study of patterns of lexis in text” (Hoey 1991: 10). Patterns of lexis are efficiently studied through corpus methods, which make it easier to identify repeated textual behaviours in large amount of data. The analysis of patterns of lexis, then, is an effective way to approach lexical cohesion from a corpus perspective. Warren (2009) exemplifies this approach analysing placement interviews in the Hong Kong Corpus of Spoken English and showing the role lexical chains play in creating cohesion in conversation. The ten most frequently occurring lexical words in the corpus are shown to be “source of lexical chains” (Warren 2009: 50), as their repetition – and the repetition of other similarly frequent and related words – across the interview establishes lexical cohesive links.

Given its contextual nature and interrelatedness, lexical cohesion is a complex issue to deal with in the translation practice. With grammatical cohesive devices (the pronoun *he*, for example), the translator is automatically aware that they function cohesively and will therefore translate them accordingly: *he* clearly refers to another referent in the text. In contrast, lexical items do not always make explicit their cohesive function, so the translator needs to take into account the whole (con)text to decide how to deal with such items. For example, *love* does not have any cohesive function *per se*, but it functions cohesively if it occurs in the same context of *affection, hate, husband*, etc. In cases like this, the translator should not only be concerned with isolated phenomena, but should also “trac[e] a *web of relationships*, the importance of individual items being determined by their relevance and function in the text” (Snell-Hornby 1988: 69, emphasis in the original). Sometimes, the translator does not or cannot reproduce the same web of relationships existing in the ST. This can create misalignment with the lexical chains of the ST and affects the cohesion of the TT. To prevent this happening, Baker (2011) suggests that:

Whatever lexical and grammatical problems are encountered in translating a text and whatever strategies are used to resolve them, a good translator will make sure that, at the end of the day, the TT displays a sufficient level of lexical and other types of cohesion in its own right. Subtle changes – and sometimes major changes – are often unavoidable. But what the translator must always avoid is the extreme case of producing what appears to be a random collection of items which do not add up to recognizable lexical chains that make sense in a given context. (Baker 2011: 216)

This is particularly crucial in the context of literary translation, where lexical chains contribute to make the fictional world cohesive. Corpus methods can be useful in this respect, in order to identify the lexical chains existing in texts, looking at frequent content words or thematically relevant items. With the help of a concordancer, it is possible to gather all the instances of a given item and examine how it is used in its context throughout the text. In Chapter 5, I will show how a corpus approach can be used to identify and study lexical chains, both in the original and in translation.

Corpus methods have been employed to study cohesion in translation, but mainly grammatical cohesion in technical translation. For instance, Krein-Kühle (2002) analyses how the English demonstrative determiner/pronoun *this*, as a cohesive device, is translated into German technical texts. The study notices that an equivalent level of coherence is maintained in the TTs, even though the translators use different and language-specific cohesive devices. Trebits (2009) focuses instead on the use of conjunctions as a way to establish cohesive links in general English texts and EU English documents, pointing out the most frequently used cohesive devices of conjunctive cohesion in the EU translations. As for the study of lexical cohesion in literary translation, the area remains mainly unexplored. One exception is Øverås's (1998) investigation of cohesive devices in the translation of fiction from Norwegian into English (and vice versa). However, the main focus of this study is testing Blum-Kulka's (1986) 'explicitation hypothesis'. My research aims therefore to fill this gap, providing a corpus-assisted study of lexical cohesion in literary translation. Specifically, Chapter 5 will prove that the interaction of units of meaning creates cohesion in the fictional representation of the African jungle. I will demonstrate that

semantic preference and semantic prosody are shared across several items, and this shared feature throughout HoD establishes a network of semantically related lexical chains. Comparing the TTs, I will also illustrate the effect of translation on these lexical networks, and how their alteration will affect the cohesion of the representation of the jungle in the Italian versions.

4.3 Units of meaning in translation

Sinclair's (1991, 2004) model considers functions spread over a unit formally larger than an individual word, with meaning resulting from the cumulative effects over several linguistic levels. This model allows a view of the linguistic form that, in a Firthian sense, is not limited to the rigid description of lexis and grammar, but also covers the pragmatic function of an item in its situational context. By taking into account this wider context it is possible to study aspects of language that exceed the formal boundaries of a text, as exemplified by the large amount of corpus research on discourse analysis (some examples are reported in Chapter 1, Section 7). These studies have demonstrated that analysing collocations, semantic preferences, and prosodies represents an effective method to gain insights into the socio-cultural implications of linguistic strategies. Translation too is concerned with socio-cultural implications, as it is inherently related to linguistic as well as non-linguistic factors. Besides being a process in which meanings and forms are shifted from one language to the other, translating also involves a situational, contextual, and cultural displacement. As I have argued in Chapter 1, the very nature of translation calls for a perspective that is able to account for both linguistic and extra-linguistic aspects, and Sinclair's (1991, 2004) model provides this perspective.

Comparing translations on the basis of the lexical item and its contextual relations makes it possible to shift the focus from word equivalence to function equivalence, as advocated in Chapter 1. The idea that meaning is not tied to the single word but is rather the result of a network of textual links puts into question the very concept of translation equivalence established at the word level. Equivalence should not be understood as a SL-word-to-TL-word relation, but instead should be studied at the level of functions enacted by the overall textual network. From a methodological

point of view, the comparison of functions across languages benefits greatly from the availability of corpus evidence, which enables the researcher to compare the formal patterning of the linguistic function both in the SL and in the TL. Tognini-Bonelli (2001: 150) argues in favour of this ‘functional equivalence’ and suggests a series of progressive steps to identify and evaluate comparability across languages adopting Sinclair’s (2004) model and a corpus approach.

The first step consists in identifying the formal patterns of a given SL lexical item and assigning the related meaning/function to each of the specific patterns, on the basis of SL corpus evidence. In the second step, a *prima facie* translation is assigned to each meaning/function that item enacts. Ideally, this step should be informed by translation corpora but, in the absence of these, the *prima facie* translation is assigned with the help of reference books, together with the past experience of the translator. The third and final step replicates the first step but in the opposite direction, moving from a TL function, as represented by the *prima facie* translation, to its formal patterning realisations in the TL. Once the relation between the SL item and the TL item has been established, it is possible to investigate their equivalence on a functional level, namely whether the item develops the same function in both languages. Any area of non-correspondence would represent an issue for the translator, whose task “is exactly that of bridging these gaps in the light of the linguistic and extra-linguistic constraints” (Tognini-Bonelli 2001: 150).

Tognini-Bonelli’s (2001) method for the comparison of units of meaning in translation is based on Sinclair’s (1991, 2004) model and, as such, it raises similar issues when applied to the specific needs of this study. For example, Tognini-Bonelli (2001) advocates the use of comparable corpora in the first and third steps as an “absolute necessity to establish equivalence” (Tognini-Bonelli 2001: 154). However, the way a lexical item is built in HoD might not be matched in any other corpus, being characteristic of HoD alone. What this study seeks to investigate is how HoD’s specific themes are conveyed through HoD’s specific formal patterns, even though these patterns cannot be exemplified through external corpus evidence. Moreover, Tognini-Bonelli’s (2001) model regards translation equivalents as comparable only when they are described as “functionally complete” units (Tognini-Bonelli 2001: 155). However, I have argued that this functional completeness is a theoretical

artefact that may not be reflected in the way a function/meaning is actually conveyed in a work of literature.

For example, consider the case of *darkness* again. As mentioned before, this word occurs only 25 times, displaying no evident collocational or colligational tendency. However, *darkness* shares its semantic preference and prosody with other related terms in the text and, with them, establishes a link that contributes to meaning. As a result, it develops an important function in the text even though it is not, as a unit, functionally complete. Following Tognini-Bonelli (2001), only functionally complete multi-word units become available as comparable units of meaning across languages; as a consequence, it would not be possible to establish a functional equivalence between *darkness* in the ST and its translations. It is therefore necessary to find an alternative model that would allow for such a comparison. In line with the analysis of the original, it makes sense to consider local textual functions as the currency for translation comparison, instead of units of meaning. The notion of local textual function not only fits better the needs of this study for the investigation of the ST, but also works equally well when it comes to comparing the TTs. It is the model with which functions in HoD have been identified, so it is also a matter of consistency to adopt it for the analysis of the translations.

The comparison of local textual functions grounds the third step of the methodology, through which Research Question 3 will be addressed, namely what effects translation has on the lexico-semantic patterns which convey the major themes in the Italian versions. The patterns identified in HoD in relation to the fictional representation of the African jungle and the African natives will be studied in the TTs as well. The aim is to investigate whether they establish the same local textual functions or not. If not, the analysis will examine the consequence of the mismatch. In this respect, this procedure aims to identify those “blank areas of no match” (Tognini-Bonelli 2001: 150) in which divergences develop, as divergences can signal a more patent intervention of the translator.

4.4 The bigger picture: Principal component analysis

As I have explained in the previous section, the third step of the methodology is concerned with the comparison of the translations. This comparison is based on the patterns identified in the original, in order to investigate how they have been reproduced in the TTs. These patterns have emerged in the first place from the study of the items foregrounded by the KW analysis. Therefore, although indirectly, the comparison of the translations builds on the KW analysis. However, KWs are only a part of the picture: quantitatively speaking, the patterns they develop represent just a minimal portion of the texts. KWs are the result of a frequency comparison, where only items with deviant frequencies are highlighted, whereas words occurring with similar frequency across different texts or corpora are ignored. These words are overlooked by the KW analysis and, consequentially, they are also overlooked during the comparison of the TTs. The aim of the fourth step of the methodology is to take into account this overlooked part of the source and target texts and use it to compare the translations. For this purpose, I will use principal component analysis (PCA) in the last part of this study.

PCA is a statistical procedure that, in corpus/computational linguistics and stylistics, is used to generate a measure of difference across texts or corpora. This measure is based on the performance of a given number of variables, usually the most frequent words (MFWs) in the texts. The frequency patterns of a pool of MFWs are compared across texts in order to see how overall similar or dissimilar these texts are when compared with each other. The basic principle behind this is that, for example, the 200 or 300 most frequent words in any text are likely to account for the vast majority of the tokens in it. Therefore, even though what we are comparing is the frequency of common words used in all texts (i.e. weak discriminators), they are able to account for most of the variation across the texts and can discriminate between them more reliably than a smaller set of infrequent words, i.e. strong discriminators (Burrows 2002b: 679).

In this thesis, PCA and KW analysis will work with different yet mutually compensating sets of words. Chapters 5 and 6 will focus on low-frequency content words (resulting from the KW analysis in Chapter 4); PCA in Chapter 7 will use mainly high-frequency function words. It is worth specifying that function words are

not always excluded from a KW analysis, though. Scott (2015a: online) explains that high-frequency function words can be KWs too, claiming that they may be indicators of the style of the text, rather than of what the text is about (see Chapter 4 for the three types of typical KWs). However, this does not affect the part PCA plays in this study. The reason is that I aim to explore the role of textual patterns as building blocks of the fictional world. Content words are more likely to relate to the fictional world than function words; therefore, I will intentionally focus on content words and exclude function words from the KW lists. As a result, function words will not be accounted for in Chapters 5 and 6, while PCA in Chapter 7 will shift the focus and compensate for what has been put aside in the previous chapters.

Consider, for example, Table 3.2 below, which shows the 50 MFWs in HoD sorted by frequency.

Table 3.2 50 MFWs in HoD

N	Word	Frequency	N	Word	Frequency
1	the	2,292	26	all	185
2	of	1,372	27	an	180
3	a	1,151	28	this	176
4	I	1,133	29	they	159
5	and	992	30	no	157
6	to	894	31	were	157
7	was	672	32	out	153
8	in	616	33	we	151
9	he	593	34	by	143
10	had	503	35	one	136
11	it	485	36	would	133
12	that	423	37	said	131
13	with	376	38	be	125
14	his	340	39	very	125
15	on	315	40	is	124
16	you	308	41	been	123
17	as	293	42	from	122
18	for	274	43	have	121
19	at	263	44	like	120
20	me	250	45	up	120
21	my	248	46	what	117
22	not	245	47	so	112
23	him	222	48	could	110
24	but	214	49	man	110
25	there	204	50	some	102

The vast majority of these items are very common function words; as such, they will be excluded from the KW lists, supposing that they appeared in such lists. In fact, the majority of them will not make part of the KW lists altogether. This means that almost all of the 50 MFWs in HoD will not be taken into consideration for the analyses in Chapters 5 and 6. However, the sum of the frequencies of these 50 MFWs accounts for over 46% of the total amount of tokens in HoD, even though they are only 0.9% of the total number of types in the text. PCA works exactly on these principles: (i) it uses very frequent and very common MFWs that account for the majority of tokens in the text; (ii) it uses these MFWs to discriminate between texts in a more efficient way than using infrequent but very text-specific words.

In the context of this study, the use of PCA will provide an alternative perspective to the relation between the TTs. PCA will use such a large portion of the total number of tokens in the texts that, statistically speaking, the analysis can be said to approximate comparison at the level of whole texts. One could object that, because function words appear in all texts independently from the author or genre, they are not particularly well suited to differentiate one text from the other. However, words do not function as discrete entities (cf. Section 4.2); on the contrary, they develop their meaning/function through the sum of the relationships they establish with each other. Function words are the markers of these relationships and, as such, they are also the markers of everything these relationships entail (McKenna et al. 1999: 152).

In conclusion, PCA will be adopted as a technique to address Research Question 4, namely to provide a picture of how the TTs relate to each other when looked at as whole texts, as opposed to the way they relate to each other when they are compared on the basis of specific linguistic patterns alone. The perspective that PCA will offer will be also useful for contextualising the analyses in Chapters 5 and 6 in a wider framework. The two methods – analysis of individual linguistic features and PCA – enable comparison on different but compensating levels: placing the findings one next to the other will show whether these multiple levels of analysis match in their depiction of the relationship between the TTs.

4.5 The software

The present work makes use of *WordSmith Tools* (Scott 2015b) and *Minitab 16* (Minitab Inc. 2014) as the main programmes for the analysis, while *Intelligent Archive* (Craig 2015) is used to generate the input for *Minitab 16*. *WordSmith Tools* is an extensively used software suite for linguistic analysis and the main tool of this study, as almost all the data used in this thesis have been retrieved using it. The 6th version has three main tools: Concord, used to generate concordance lines for a specific search item; WordList, which generates frequency lists of words and clusters; and Keywords, for comparing two different word lists and generating KW lists. These three main applications provide additional features to develop a great variety of tasks, giving the researcher the freedom to set many parameters and options according to the needs of the study.

In Chapter 7, I will introduce *Minitab 16*, as *WordSmith Tools* does not have any function for carrying out multivariate analyses. *Minitab 16* is a statistics package for numerical analysis that, among many other functions, allows the user to perform a wide range of multivariate analyses, including PCA. The software is able to import data from a spreadsheet and use them to produce the desired calculation. Moreover, it also has graphic functions to create, among other graphs, score plots and scatterplots (see Chapter 7 for more details). In the same chapter, I will also use *Intelligent Archive*. *Intelligent Archive* is an application developed at the University of Newcastle, Australia, specifically designed for preparing a set of data to be exported to an external spreadsheet or statistics programme such as *Minitab 16*. As stated on the application webpage, while most of the text-processing programmes focus on more linguistic outputs, such as concordance lines, *Intelligent Archive* is more statistical and only produces frequency word counts. Within the aims of this study, *Intelligent Archive* will be used to create automatically parallel wordlists.

5. Conclusion

This chapter has discussed the methodological framework that underpins the present study and has described the steps that will be carried out to answer to the four

research questions. Each step has been designed to add a new layer of scrutiny to the analysis, resulting in a progression: from the whole texts to specific linguistic features and then back from linguistic features to the whole texts (cf. Figure 3.1). The KW analysis will foreground words with frequencies that are statistically more significant than others, making them stand out in the text as a whole. Mahlberg & McIntyre's (2011) model of KW categorisation will help to recognise the major themes in these prominent words, so that a connection between lexis and content can be established. Then, these items will be analysed in detail, as developing local textual functions in the process of themes construction. The outcome of this in-depth analysis will become the starting point for the comparison of the TTs, based on functional equivalence so that extra-linguistic aspects as well as the linguistic ones will be taken into consideration. Finally, PCA will go back to the whole texts, establishing a comparison from a wider point of view and counterbalancing the local perspective the previous comparisons have provided.

Rather than delineating a linear procedure, these steps can in fact be seen as forming a methodological cycle, in which the last stage can feed in and go back where the first stage started. The analysis moves from the whole texts to individual features through the study of KWs and lexical items, then from individual features back to the whole texts through PCA, as showed in Figure 3.3 below.

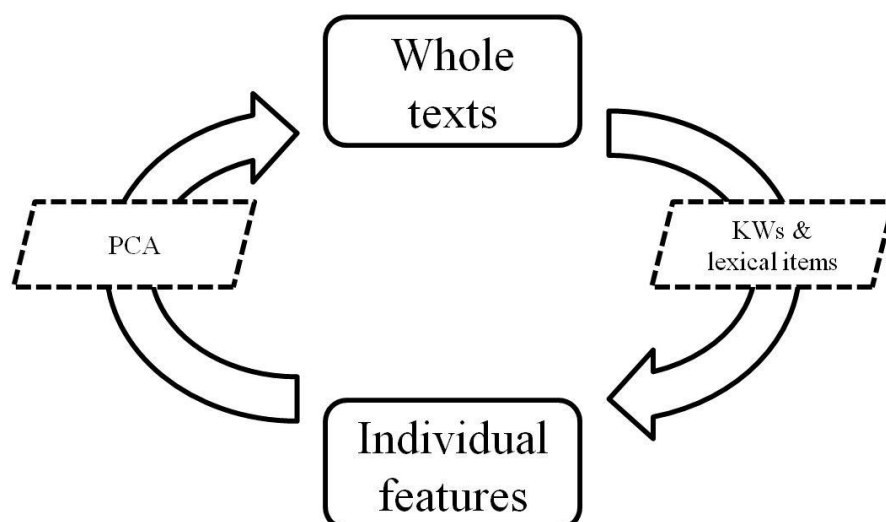


Figure 3.3 The methodological circle

Although this study will complete only one round of the circle, it is possible to start again, as the findings of PCA on the relation of the whole texts can inform the analysis of the specific lexical features. I will in fact show that the outcome of Chapters 5 and 6 and the outcome of Chapter 7 are compatible with each other, even though they are based on different levels of linguistic scrutiny. With this awareness, it would have been possible to start another circle and explore how these different layers of the study interact with each other. This approach aims to offer a more holistic picture of how literary texts and their translation work, where diverse linguistic levels are regarded as intertwined and affecting each other.

CHAPTER FOUR

From lexis to themes: The keyword analysis

1. Introduction

The first step of this study is to identify text-specific words that can be related to the representation of Africa and to the representation of the African natives in HoD. I have argued in the previous chapter that KW analysis is an optimal way to do so, as this method offers several practical advantages. It allows me to start the investigation with a controlled amount of items that, at the same time, are also statistically significant and automatically generated (cf. Chapter 3, Section 4.1). More importantly, KW analysis enables me to encompass both a text-based and a reader-based perspective to the analysis of HoD. On the one hand, KW analysis is strictly text-based. This is because KWs are retrieved on the basis of textual features only, while the reader's understanding of the text does not play a major role in their identification. KWs are characteristic items of a text, as their frequency is significantly higher or lower compared to the norm. They are generated blindly, only on the basis of their frequencies, without reflecting any "meaning relationships between them" (Mahlberg & McIntyre 2011: 206). On the other hand, grouping KWs into thematic categories reflects a specific understanding of the text, rather than formal features. As such, the categorisation of KWs provides a reader-based perspective. As this study aims to investigate the themes of Africa and its representation and race and racism, KWs will be categorised in accordance with this aim. These two themes have emerged from the review of the literary criticism (Chapter 2), which can be considered an informed, shared, and consolidated understanding of the text. Finally, generating KWs and grouping them into thematic categories represents a way to link text-specific features to a reader-based understanding of the text, that is, that of the critics. This link, once again, is in line with the philological circle underpinning stylistic analyses (Chapter 1), which seeks to relate linguistic description and literary appreciation.

In Section 2, I will describe the procedure to generate the two KW lists used in this study. I will also discuss methodological aspects of the KW analysis, such as the

selection of a reference corpus and the statistical cut-off point used to generate KWs, together with their implications. Then, comparing the two lists, I will demonstrate the advantages of using multiple reference corpora. In Section 3, I will concentrate on Mahlberg & McIntyre's (2011) model for the classification of KWs, arguing that this particular model fits the needs of the present study. In Section 4, I will offer an overview of the KW classification and describe some of the categories identified, while, in Section 5, I will explain which categories and KWs are selected to be studied in detail in the following chapters. I will introduce the resulting "Africa words" and "African words", discussing the items they contain and the reasons behind their selection. Finally, Section 6 will close the chapter with some brief concluding remarks.

2. Generating the KW lists

In a KW analysis, the choice of which reference corpus to use to generate KWs is a fundamental methodological decision. This decision can in fact affect the outcome of the analysis, as using different reference corpora usually produces different sets of KWs. In order to minimise the effects of choosing a reference corpus and make the analysis more reliable and empirical, I will use multiple reference corpora. Specifically, two lists will be generated: the first, using CC; the second, using RC as a reference corpus. I will also compare the resulting KW lists, as I aim to identify words that are as specific to HoD as possible. Fischer-Starcke (2010) explains that generating multiple sets of KWs by comparing a text to different reference corpora enables the comparison of the resulting sets. Words that are shared among lists are more significant, as their status as a KW does not depend on any individual reference corpus: being key in more than one set of data makes KWs better indicators of the distinctive aspects of a text.

The first list (List 1) is generated by using CC as a reference corpus. CC is adopted as representative of Conrad's language-internal norm, so that comparing HoD to CC will identify those words that characterise HoD against Conrad's fictional language. The second list (List 2) is obtained by employing RC as a reference corpus. As RC seeks to represent the external norm of the literary language of the period,

comparing HoD to RC will emphasise those items that characterise Conrad's short novel against general fictional language. Comparing the two KW lists will foreground items that are fundamentally typical of HoD. Only those words that are shared by both lists will be selected, in this way minimising the risk that they could reflect Conrad's linguistic idiosyncrasies, rather than intrinsic features of HoD. At the same time, this comparison ensures that the resulting KW categories reflect themes that are HoD-specific, as opposed to themes that are generally central in Conrad's works.

The selection of the reference corpus is not the only procedural decision that has a major effect on the outcome of the analysis. Another criterion that affects significantly a KW analysis is the statistical cut-off point for considering a word a KW. As explained in Chapter 3, KWs are generated by comparing the relative frequency of the words in a text with the relative frequency of the same words in a reference corpus. If the frequency of a given word in the text is unusually high compared to its frequency in the reference corpus, then that word is considered a KW. Programmes such as *WordSmith Tools* use statistical tests – usually chi-square or log-likelihood – to check whether the difference in frequency is statistically significant or rather due to chance. If the result of the statistical test is higher than a given cut-off point, then it means that the difference in frequency is statistically significant; as a result, that word will be considered a KW. However, this cut-off point is not fixed but can be set by the user of the software. The decision is up to the researcher and different cut-off points produce different sets of KWs.

Consider log-likelihood and the p -value, which are used in this analysis to generate KWs with *WordSmith Tools*. The p -value threshold represents the cut-off point that the software adopts to define a KW. *WordSmith Tools* uses a log-likelihood test to check whether the frequency difference between the word in our text and the same word in the reference corpus is statistically significant. If the p -value resulting from the comparison is lower than the cut-off point, then that word in our text is recognised as a KW. A p -value of 0.05, for example, indicates a 95% confidence that the outcome of the KW analysis is not due to chance; a p -value of 0.01 indicates a 99% confidence; a p -value of 0.0001 indicates a 99.99% confidence; and so on. The lower one sets the p -value, the fewer KWs one obtains, since the frequency difference must be more striking (i.e. more statistically significant) to pass the cut-off point. The

selection of the p -value threshold therefore has important implications for the analysis.

However, there is no agreement over the “right” cut-off point for a KW analysis (Baker 2010: 26) and the selection of which threshold to use may depend on the needs of the study. Setting too high a p -value threshold usually generate too many KWs. McEnery (2009: 98) explains that, while it is indeed possible to analyse a large number of KWs, a smaller and more coherent sub-set is preferable (as a reference, Baker (2006: 127) considers 22 KWs a “manageable” amount, although what is “manageable” and what is not depends on the type of analysis). At the same time, too low a p -value cut-off point can exclude important words from the investigation, especially when the text under analysis is not particularly large. Baker (2010: 26) points out that the larger the texts or corpora we are examining, the larger the amount of KWs we are likely to obtain. Given that HoD is only about 39,000 words long, by setting a very low threshold we might run the risk of cutting out relevant words related to the representation of Africa and of the natives. For example, comparing HoD with CC by setting a p -value of 0.0001 generates 88 KWs, while lowering the p -value to 0.0000001 (which is *WordSmith Tools* default setting), generates only 28 KWs, cutting out words such as *savages*, *darkness*, and *niggers* which I will show to be relevant for the analysis of the representation of the African jungle and the African natives. I will therefore use as cut-off point a p -value of 0.0001, which indicates 99.99% confidence that the results of the KW analysis are not due to chance. In this way, the analysis is not penalised by the fact that HoD is a short text. Setting a p -value of 0.0001 and a minimum frequency of 3 generates 88 KWs for List 1 (HoD vs. CC) and 183 KWs for List 2 (HoD vs. RC).

Once the two lists have been generated, all function words are deleted, so that only content KWs remain. There are two reasons behind this decision. First, excluding function words reduces further the number of KWs to consider in the analysis. This is particularly important, given that combining the two lists without removing function words would produce 271 KWs, which are too many to analyse in detail. Most importantly, content words are more likely to reflect major themes. Even though function KWs have been shown to be equally indicative of stylistic features (Culpeper 2002b, 2009), this study aims specifically to establish a link between the lexical level of HoD and its thematic concerns. I argue that this link is better

established by looking at content words rather than function words, because content words relate directly to the “aboutness” of the text (Scott 2015a: online) and “tell us what a text is about” (Stubbs 2002: 39) by conveying its semantic content. Removing function words reduces the number of KWs to 76 in List 1 and to 163 in List 2. The two lists are reported in Table 4.1 and Table 4.2 below.

Table 4.1 List 1 (HoD vs. CC; content words only)

N	Keyword	N	Keyword	N	Keyword	N	Keyword
1	kurtz	20	earth	39	intrusted	58	patches
2	manager	21	cipher	40	caravan	59	screech
3	pilgrims	22	staves	41	fresleven	60	lots
4	station	23	pilot	42	etc	61	arrows
5	kurtz’s	24	steamer	43	wheel	62	work
6	ivory	25	brooding	44	rifle	63	administration
7	steamboat	26	see	45	surf	64	science
8	rivets	27	stream	46	banks	65	tin
9	river	28	savages	47	restraint	66	ominous
10	wilderness	29	waterway	48	cutters	67	chaps
11	snag	30	fossil	49	months	68	crawled
12	bush	31	nightmares	50	sometimes	69	introducing
13	shutter	32	woods	51	eloquence	70	kind
14	bank	33	grass	52	helmsman	71	mend
15	forest	34	wool	53	attack	72	hens
16	boiler	35	reaches	54	darkness	73	things
17	pole	36	manager’s	55	invasion	74	agent
18	hippo	37	unsound	56	knights	75	niggers
19	stretcher	38	martini	57	shamefully	76	fog

Table 4.2 List 2 (HoD vs. RC; content words only)

N	Keyword	N	Keyword	N	Keyword	N	Keyword
1	kurtz	42	very	83	tin	124	murmur
2	manager	43	screech	84	swayed	125	yells
3	pilgrims	44	pole	85	rifle	126	villages
4	kurtz's	45	black	86	aspect	127	nowhere
5	river	46	patches	87	ashore	128	absurd
6	ivory	47	trading	88	staves	129	imagine
7	steamboat	48	cutters	89	too	130	steam
8	station	49	nightmares	90	sometimes	131	tobacco
9	rivets	50	administration	91	yarns	132	row
10	wilderness	51	darkness	92	carriers	133	glitter
11	steamer	52	reaches	93	russian	134	frightful
12	shutter	53	immensity	94	europe	135	lank
13	forest	54	etc	95	below	136	months
14	snag	55	banks	96	devil	137	warlike
15	marlow	56	agent	97	crawled	138	approach
16	bush	57	leaped	98	depths	139	bearers
17	earth	58	martini	99	desolation	140	canoes
18	coast	59	winchesters	100	immense	141	gleams
19	helmsman	60	fresleven	101	fog	142	wool
20	deck	61	waterway	102	indistinct	143	cliff
21	bank	62	ships	103	feet	144	hanged
22	pilot	63	mud	104	intrusted	145	heard
23	shore	64	grass	105	fireman	146	cipher
24	boiler	65	brooding	106	fusillade	147	hut
25	stretcher	66	nigger	107	whites	148	continent
26	lot	67	riverside	108	rags	149	gloom
27	hippo	68	ominous	109	seemed	150	myself
28	seaman	69	fellows	110	sea	151	steamed
29	sombre	70	confounded	111	whisper	152	attack
30	trade	71	chap	112	pitiless	153	miles
31	savages	72	glance	113	overboard	154	sorrow
32	stream	73	impenetrable	114	mystery	155	fright
33	creek	74	abreast	115	introducing	156	arrows
34	surf	75	eloquence	116	shamefully	157	fringed
35	niggers	76	unsound	117	stillness	158	see
36	restraint	77	invasion	118	inconceivable	159	savagery
37	wheel	78	profound	119	lost	160	hammock
38	savage	79	lots	120	purpose	161	aft
39	chaps	80	swede	121	amazing	162	starched
40	fossil	81	afar	122	bends	163	cotton
41	company's	82	manager's	123	current		

Comparing the two lists, it can be observed that the vast majority of the items in List 1 appear in List 2 as well. This suggests that what distinguishes HoD among Conrad's works also characterises it against other fictional works of the same period. Besides proper names (which are text-specific in most of cases), KWs such as *ivory*, *river*, *darkness*, *forest*, *wilderness*, *niggers*, *station*, and *steamboat* indicate the specific content of the short novel and distinguish it against other works by both Conrad and other writers. As suggested by Fischer-Starcke (2010: 66), the fact that they appear in both lists maximises their significance as characteristic features of HoD, bringing support to their examination as potential indicator of HoD-specific themes.

Looking at those words that do not appear in both tables is equally indicative of the advantages of using two reference corpora. For instance, if only List 1 had been considered, it would have seemed that *science* and *work* are typical of HoD. However, this would have been misleading. List 2 reveals that these words occur with a comparable frequency in other texts of the same period, suggesting that they are not as specific to HoD as it would have appeared if only one reference corpus was used. Focusing on the KWs that appear in List 2 but not in List 1 shows that more than half of the words in List 2 are not shared by the other list. This means that if I had compared HoD to RC only, I would have had more than 60 KWs that are not as specific to HoD as I originally wanted, but that rather represent lexical preferences or idiosyncrasies of Conrad's overall fictional language.

Consider *sea* and *ships*, for example, which appear in List 2 but not in List 1. HoD is about a journey by water and these two KWs seem to reflect this topic. However, HoD is more specifically about navigation along a river. Seafaring and ships are certainly common themes in Conrad's fiction, but they do not single HoD out among Conrad's other novels. Therefore, *sea* and *ships* do not appear in List 1. Another example is provided by *creek* and *riverside*. More than *sea* and *ships*, these two KWs seem to refer precisely to the journey on the Congo River and as such appear to reflect specifically the content of HoD. Yet, KWs such as *waterway* or *bank* would reflect more characteristically this topic, because they are shared by both lists. Thus, comparing the two lists tells us that *waterway* and *bank* are better candidates to analyse in detail than *creek* and *riverside*.

3. The KW classification model

The comparison of List 1 to List 2 in order to select those items that appear in both lists produces a final list of 67 KWs. These KWs can be considered specific of HoD; I will use them as the starting point for the creation of the categories. The final list of KW is presented in Table 4.3 below.

Table 4.3 KWs shared by List 1 and List 2

N	Keyword	N	Keyword	N	Keyword	N	Keyword
1	kurtz	18	hippo	35	manager's	52	darkness
2	manager	19	stretcher	36	unsound	53	invasion
3	pilgrims	20	earth	37	martini	54	shamefully
4	station	21	cipher	38	intrusted	55	patches
5	kurtz's	22	staves	39	fresleven	56	screech
6	ivory	23	pilot	40	etc	57	lots
7	steamboat	24	steamer	41	wheel	58	arrows
8	rivets	25	brooding	42	rifle	59	administration
9	river	26	see	43	surf	60	tin
10	wilderness	27	stream	44	banks	61	ominous
11	snag	28	savages	45	restraint	62	chaps
12	bush	29	waterway	46	cutters	63	crawled
13	shutter	30	fossil	47	months	64	introducing
14	bank	31	nightmares	48	sometimes	65	agent
15	forest	32	grass	49	eloquence	66	niggers
16	boiler	33	wool	50	helmsman	67	fog
17	pole	34	reaches	51	attack		

Scott (2015a: online) explains that there are three types of words that usually make it into a KW list. The first type is “proper nouns”. In Table 3, this type is represented by *Kurtz*, *Kurtz's*, and *Fresleven*. The second type is “aboutness KWs”, that is, those words that give a good indication of the content of a text. This is the most relevant type of KW for this study because, as I have argued in the previous section, these words enable me to establish a link between the lexical level of HoD and its major themes. The vast majority of the entries in Table 4.3 belong to this type, for example *bank*, *administration*, *river*, *ivory*, *pilgrims*, *station*, *darkness*, *steamboat*, *niggers*, and *wilderness*. The third type of KW is represented by high-frequency words that may be indicators of style rather than of “aboutness”. This type of KWs is underrepresented in Table 3, as function words have been removed from the two original lists. However,

even though not high-frequency, there are some words in the table that are more likely to be related to the style of the text rather than to its content. For instance, KWs like *etc* or *unsound* are certainly not signals of what HoD is about, but rather seem to indicate some of Conrad's stylistic preferences, such as the use of negative prefixes (Stubbs 2005: 16) or the use of *etc* to shorten a dialogue line that goes uselessly on for too long.

Scott's (2015a: online) "aboutness" KWs include exactly the type of KWs that are relevant to this study, namely those that reflect the content of the text. However, this classification is too general for the aims of this work, as it does not make any further distinction of the different aspects and levels within the content of a text.⁶ Limiting the grouping of KWs to this category only would result in a large and undifferentiated amount of content words, too general to refer to specific aspects of the fictional world, such as the African jungle and the African natives. A further level of discrimination is therefore required. More complex models for the classification of KWs have been discussed in corpus linguistic literature (cf. Chapter 3, Section 4.1) but they have not been designed to be applied to literary texts and may not work equally well with HoD. Thus, I will adopt Mahlberg & McIntyre's (2011) model of KWs classification, which better fits the needs of my analysis.

Mahlberg & McIntyre's (2011) model was originally designed to study KWs in Ian Fleming's *Casino Royale* as "signals for the building of fictional worlds as well as triggers for thematic concerns of the novel" (Mahlberg & McIntyre 2011: 207). Mahlberg & McIntyre (2011) classify KWs into two main categories: 'fictional world signals' and 'thematic signals'. Fictional world KWs are defined as "world-building element[s]" (Mahlberg & McIntyre 2011: 209) that, with their concrete meaning, contribute to establishing the fictional environment of the novel. For example, Mahlberg & McIntyre (2011: 209) classify *table* as a fictional world KW, because it has a concrete meaning that refers specifically to the fictional world of the novel: *table* is used repeatedly to refer to the gaming tables in the casino. In the case of HoD, *bush* and *station* are two examples of fictional world KWs, as they contribute to establish the fictional setting in which the short novel is set. Thematic signal KWs, on the other hand, are less concrete than fictional world KWs: they can be ambiguous,

⁶ Scott's (2015a: online) three types of KWs are in fact just meant to describe the most common kinds of KWs, rather than offering a classification model.

metaphorical, or evaluative because of their polysemy. As such, they are open to wider interpretations (Mahlberg & McIntyre 2011: 209). These KWs can signal themes of the text and work in conjunction with other textual items – for example, other fictional world and/or thematic signal KWs – in order to establish these themes. Mahlberg & McIntyre (2011: 209) identify *gambler* as a thematic signal KW in *Casino Royale*. Besides its concrete meaning related to the casino setting, *gambler* can also be interpreted as referring to a “risk-taker” attitude, embodied by James Bond. Together with other KWs, *gambler* can be seen as establishing the theme “taking risks” (Mahlberg & McIntyre 2011: 209). In HoD, an example of thematic signal KW is *wilderness* since, in addition to its concrete meaning of “wild land covered in vegetation”, it can also be considered as the opposing force that resists the Western colonisers and be interpreted as signalling broadly the theme “colonisation and its consequences”.

Fictional world signal and thematic signal KWs are the two main categories of the model. These two categories are further broken down into sub-categories that group more specifically different aspects of the fictional world or diverse themes of the text. For instance, Mahlberg & McIntyre’s (2011) fictional world category is divided into “characters” and “settings and props” sub-categories, each of them with its own sub-groups. This system of sub-categories allows me to pin down exactly the aspects of HoD I am interested in, and to populate the relative sub-groups with KWs to be analysed in depth in the successive stages of the analysis.

As I have mentioned in Section 1, the value of grouping KWs for a corpus stylistic study also derives from the fact that these categories enable me to take into account critical interpretations. Categories are shaped in accordance to a subjective understanding of the text which, in the case of the present study, is represented by the established interpretations of the literary critics. By populating these categories, then, I will create a link between the text-based features of HoD and their interpretation, although to different extents. Fictional world KWs are in fact more text-based than reader-based, in the sense that they reflect more closely lexically-driven features of the text, rather than their subjective understanding by the reader. A character’s name (*Kurtz*) or a descriptive reference to the novel’s setting (*river*) are less open to interpretation than, say, a polysemous term such as *wilderness* or *darkness*. Thematic

signal KWs are instead more reader-based than text-based, as they work on multiple levels of meaning and require more interpretative efforts.

However, this does not mean that only thematic signal KWs are able to account for the interpretation of HoD. Even though fictional world KWs refer to more concrete aspects of the text and are therefore less susceptible to subjective interpretations, they can equally have interpretative implications when contextualised. Let us consider *forest*, for example. At first, this KW can appear neutral or concrete, as it clearly refers to the setting of HoD, the Congolese forest. As such, *forest* is classified as fictional world signal KW. Nonetheless, I argue that the fictional world plays a fundamental role as trigger of thematic concerns in HoD: as I have shown in Chapter 2, the theme “Africa and its representation” is strictly related to the way the jungle is depicted in the short novel. *Forest*, then, participates in the building of the theme of Africa and, in conjunction with other fictional world KWs, accounts for a more reader-based aspect of the text. In the two following chapters, I will demonstrate how fictional world signal KWs are related to the two major themes emerged from the discussion in Chapter 2. I will bring evidence to the argument that KW categories – fictional world signal and thematic signal – can be seen as lexically-driven features of the text where both text-based and reader-based perspectives interact.

4. The KW categories

In order to create the categories, I have examined each item in the final list of KWs (Table 4.3) with the help of the concordancer; words with related meaning have been put together into the same group. Each successive KW that does not fit in one of the existing groups leads to the creation of a new group, until the exhaustion of the items. In this process, the categories have been reshaped and adjusted so as to obtain the best fit for the KWs, in line with the aim of the model that is establishing the categories themselves, rather than quantifying the KWs (cf. Mahlberg & McIntyre 2011: 209). Table 4.4 below shows the resulting categories.

Table 4.4 ‘Fictional world signal’ and ‘thematic signal’ KWs

Category	Keywords
<i>Fictional world signal</i>	
<i>Characters</i>	
Names	Kurtz, Kurtz’s, Fresleven
Equipment & tools	pole, stretcher, Martini, rivets, staves, wool, rifle, arrows, tin, hippo
Natives	savages, cutters, chaps, niggers, helmsman
Colonisers	manager, manager’s, pilgrims, chaps, agent
<i>Setting & props</i>	
Places	earth
Station	manager, manager’s, pilgrims, chaps, administration, station, agent, ivory, fossil
Forest	bush, forest, hippo, grass
River	river, stream, waterway, reaches, surf, bank, banks, snag
Boat & sailing	steamboat, boiler, steamer, pilot, wheel, screech, crawled, shutter
<i>Temporal indicators</i>	months, sometimes
<i>Atmosphere</i>	
Negativity	brooding, nightmares, restraint, ominous, darkness
<i>Thematic signal</i>	wilderness, darkness, invasion, fog
<i>Unclassified</i>	etc, eloquence, lots, cipher, see, intrusted, attack, shamefully, introducing, unsound, patches

Table 4.4 shows that all the KWs resulting from the comparison of List 1 and List 2 (Table 4.3 above) have been classified in the two main categories of the model, fictional world signal and thematic signal KWs, plus an extra “Unclassified” group. This group collects those items that do not fit in any of the existing sub-categories. This does not necessarily mean that these words are not significant, but simply that there are not enough KWs to create separate categories with them. If more KWs were taken into account, perhaps the items in the “Unclassified” category might have formed further sub-categories of their own. As it is, they will not be considered in this study.

The fictional world signal KWs are divided into four sub-categories, representing fundamental aspects of the process of building up the fictional world.

These sub-categories are “Characters”, “Setting & props”, “Temporal indicators”, and “Atmosphere”. It is interesting to notice that the “Characters” sub-category is very different from that obtained by Mahlberg & McIntyre (2011), a difference that is indicative of specific aspects of HoD. First of all, the sub-group “Names” contains two names only, *Kurtz* and *Fresleven*. Considering that the latter is a very minor character in the story, it can be said that “Names” is not a very populated category in HoD. In contrast, Mahlberg & McIntyre (2011) have six names in the same sub-group, and they even specify that their “Names” group only includes major characters (Mahlberg & McIntyre 2011: 211). In HoD, characters are rather referred to with common nouns, as indicated by the sub-groups “Natives” and “Colonisers” which are more largely populated than “Names”. This is not only the case for general groups of people (*savages*, *niggers*, *cutters*, and *chaps* for the natives; *pilgrims* and *chaps* for the colonisers), but it happens also with specific characters that have more than a passing role in the story, such as the African helmsman of Marlow’s steamboat or the manager of the station. The paucity of proper nouns and the abundance of common nouns in the “Characters” sub-category confirm Stubbs’s (2005: 8) observation about HoD that “people are not named, but identified by their functions”. More generally, this relates to discussion about the text’s vagueness, a feature that many critics have recognised. Leavis (173: 180), for example, has famously argued that Conrad makes a virtue out of not knowing what he means, Murfin (1989: 131) claims that HoD is a “masterpiece of concealment”, while O’Pray (1983: 23-24) recognises a “climate of doubt and vagueness” that surrounds and characterises HoD.

This vagueness is paralleled in the “Setting & props” and in “Temporal indicators” sub-categories. Stubbs (2005: 8) argues that HoD is full of “examples of this vagueness about places, times, and people”, specifying that major places are never named and that the timescale of the story is similarly very vague. The “Temporal indicators” sub-category includes two KWs: *months* and *sometimes*. *Months* occurs 17 times, but only 4 times it is preceded by a precise number (*nine months*, *six months*, *three months*, *two months*). The rest of the time, it occurs with a more abstract adjective – mainly *some months* (3 times), *few months*, and *several months* (2 times) – or without any specification at all, as in *for months* (3 times) and *only months and months afterwards*. The frequent occurrence of *sometimes* (24 times) reinforces this sense of indefinite and unstated time that permeates the text. The

insistent recurrence of *sometimes* is part of a larger stylistic feature: Stubbs (2005: 10) shows that more than 200 occurrences of *something*, *sometimes*, *somewhere*, *somehow*, and *some*, together with other expressions of vagueness such as *kind of* and *sort of*, occur in HoD.

The “Setting & props” sub-category mirrors the “Characters” sub-category because it does not include any proper nouns. Proper place nouns are common in other KW analyses of literary texts. For example, the “Places” sub-group in Mahlberg & McIntyre (2011)’s study includes *Royal-les-Eaux* and *Paris*, while Fischer-Starcke’s (2009: 521) study of Austen’s *Pride and Prejudice* identifies, among the others, *Meryton*, *Hertfordshire*, *Netherfield*, and *Longbourn*. The absence of proper place nouns in HoD’s KWs is revealing, especially considering that proper names are generally regarded as one of the most frequent types of KWs in any text (Scott 2015a: online). Conversely, major places in HoD are mostly referred to with general common nouns, as exemplified by *station*, *forest*, and *river*, three of the KWs with the highest keyness value that relate to the principal locations of the story.

The “Atmosphere” sub-category is not included in Mahlberg & McIntyre’s (2011) original classification and reflects a different aspect of the text compared to “Characters”, “Setting & props”, and “Temporal indicators”. The items in “Atmosphere” are not as concrete as those in the other fictional world signal sub-categories, but they equally shape the fictional world, although in a different way. Instead of relating to specific characters or places, words such as *brooding*, *nightmares*, *unsound*, *restraint*, *ominous*, and *darkness* strengthen that negative and gloomy feeling, that dark atmosphere that lies heavily on HoD and characterises it. They still contribute to building the fictional world, but emotionally rather than materially. Consider *brooding* and *ominous*, for example, whose concordance lines are shown in Figure 4.1 and Figure 4.2 below.

1	the place of the monstrous town was still marked ominously on the sky, a brooding gloom in sunshine, a lurid glare under the stars. "And this also,"
2	his work was not out there in the luminous estuary, but behind him, within the brooding gloom. Between us there was, as I have already said somewhere,
3	and that frown gave to his black death-mask an inconceivably sombre, brooding, and menacing expression. The lustre of inquiring glance faded
4	starvation, its exasperating torment, its black thoughts, its sombre and brooding ferocity? Well, I do. It takes a man all his inborn strength to fight
5	Gravesend, and farther back still seemed condensed into a mournful gloom, brooding motionless over the biggest, and the greatest, town on earth. The
6	as if about to go out suddenly, stricken to death by the touch of that gloom brooding over a crowd of men. Forthwith a change came over the waters, and
7	not in the least resemble a peace. It was the stillness of an implacable force brooding over an inscrutable intention. It looked at you with a vengeful aspect.
8	stood looking at us without a stir, and like the wilderness itself, with an air of brooding over an inscrutable purpose. A whole minute passed, and then she
9	and draping the low shores in diaphanous folds. Only the gloom to the west, brooding over the upper reaches, became more sombre every minute, as if

Figure 4.1 Concordance lines of *brooding*

1	uneasy. You know I am not used to such ceremonies, and there was something ominous in the atmosphere. It was just as though I had been let into some
2	the patient woods, those broken phrases came back to me, were heard again in their ominous and terrifying simplicity. I remembered his abject pleading, his abject threats,
3	about for I can't guess. Embalm it, maybe. But I had also heard another, and a very ominous, murmur on the deck below. My friends the wood-cutters were likewise
4	her. She was savage and superb, wild-eyed and magnificent; there was something ominous and stately in her deliberate progress. And in the hush that had fallen
5	The opening paragraph, however, in the light of later information, strikes me now as ominous. He began with the argument that we whites, from the point of development
6	that come to one sometimes. The high stillness confronted these two figures with its ominous patience, waiting for the passing away of a fantastic invasion. "They swore
7	suddenly of that ship of war I had seen firing into a continent. It was the same kind of ominous voice; but these men could by no stretch of imagination be called enemies.

Figure 4.2 Concordance lines of *ominous*

Followed by *over 5* out of 9 times, *brooding* suggests the presence of a bleak aura, a negative atmosphere weighting down on the events and characters of the book. *Brooding* also occurs with other equally negative terms, especially *gloom* (5 times) and *sombre* (2 times). *Ominous* similarly participates in this effect, emphasising the mysterious quality of everything in the story. *Voice* (line 7) and *murmur* (line 3) are *ominous*, as well as the *patience* (line 6) of the African forest and the appearance of the native woman that Marlow meets in one of the stations (line 4). Revealingly, in line 1 it is the *atmosphere* that has something *ominous* about it.

Generally speaking, “Atmosphere” KWs can be considered halfway between fictional world signal and the thematic signal KWs. On the one hand, as I have shown with these examples, “Atmosphere” KWs contribute to shaping the fictional world by describing aspects and defining details of the characters, places, and events, not in a concrete sense but rather through evoking atmosphere. In this respect, they act as fictional world KWs. On the other hand, their abstract nature enables them to convey wider or multiple meanings, and therefore reflect broader themes. Critics have argued that one of the aims of Conrad with HoD was to throw over the reader “a brooding gloom”, to convey a “resonance and tenebrous atmosphere” (Guerard 1966: 44). Therefore, “Atmosphere” KWs in HoD can also act as thematic signal KWs. It is not a case that *darkness* appears in both the “Atmosphere” sub-category and in the thematic signal category: I will discuss the function and textual patterns of *darkness* in detail in the next chapter, showing how it contributes to the theme of Africa and its representation.

Thematic signal KWs are not grouped into further categories, because they are much less numerous than fictional world signal KWs. However, even if they were more numerous, it would still be difficult to group them into sub-categories, as these words convey different meanings and work on multiple levels at the same time, as a result of their polysemy. As such, it would not be easy to pin them down into any

specific group reflecting a given use. This seems to be an intrinsic characteristic of thematic signal KWs: their polysemy, Mahlberg & McIntyre (2011: 212) explain, “is relevant to viewing words as thematic signals, which relates to the fact that polysemy is a concept strongly associated with literariness”. Similarly to my classification, Mahlberg & McIntyre’s (2011: 210) KW grouping identifies only 5 thematic signal KWs, with no further subdivision.

As mentioned in Section 2, *wilderness* can be seen as conveying different meanings. It has a concrete meaning of “wild land covered in vegetation”, but it can also be interpreted as the inner spirit of the African forest, the contrasting force that resists the colonisation process. *Darkness* too works on multiple levels. Besides its concrete meaning of “absence of light”, the concept behind this word has been interpreted in several different ways. *Darkness* can refer to the dark centre of the African forest, in the literal sense of “heart of darkness”, but it can also refer to the dark heart of the imperialistic enterprise. Both *wilderness* and *darkness* play a fundamental role in the representation of Africa and will be discussed in detail in the next chapter.

The KW *invasion* similarly develops a twofold meaning. First, *invasion* refers to the concrete presence of a given character or group of characters in a place where they should not be: for example, the presence of Kurtz in the depths of the wilderness is defined as *invasion*. At the same time, *invasion* can also be interpreted as referring to the entire colonial mission, and to the overall imposition of the Western culture forced on the non-Western world. In fact, in 4 out of 5 cases, it is the African natives or the wilderness itself that wait patiently for *the passing away of this fantastic invasion*. This is far from saying that this pattern alone conveys the theme of “invasion”. However, in conjunction with other items (maybe other KWs), *invasion* may well be seen as contributing to the establishment of this theme.

This is equally valid for *fog* which, more than *invasion*, possesses a concrete meaning. However, *fog* does not refer to any specific place in particular (and consequentially it is not simply a fictional world signal KW), but its presence is spread out across the text. Watts (2006: 350) argues that mist and haze are persisting images in HoD, which express the fugitive and indefinite nature of the impressionist experience the text enacts. The haze that surrounds Marlow’s story symbolises the ambiguous quality of any individual experience: “*Heart of Darkness* embodies more

thoroughly than any previous fiction the posture of uncertainty and doubt” (Watts 2006: 355). It can be said that the theme “uncertainty of individual experience” (which is related to the theme “vagueness” discussed above) is reinforced by the pervasive use of the fog/mist/haze image. Stubbs (2005) points out that *fog* and other semantically related words (*blurred, dusk, mist/misty, murky, shade, smoke, dark, etc.*) are used almost 150 times in the text: “Marlow is often looking into a fog, uncertain of what he is seeing” (Stubbs 2005: 9).

5. The “Africa words” and the “African words”

In the previous section, I have offered an overview of the KW categories, illustrating how they touch upon a wide range of different aspects of the fictional world and different themes. However, my research will focus on two specific aspects of the fictional world, aspects that are related to the two major themes emerged from the discussion in Chapter 2, “Africa and its representation” and “race and racism”. I have argued that these two themes are strictly linked to the way Africa and the Africans are fictionally represented in the text. I will therefore concentrate on categories and KWs that refer directly to the African forest and the natives. This focus will result in the two groups of words, which will be analysed in detail in the following chapters. In the next sections, I will introduce them and explain the criteria beyond their selection. In Section 5.1, I will introduce the “Africa words”, while in Section 5.2 I will present the “African words”.

5.1 The “Africa words”

The obvious starting point for the discussion of the representation of the African forest is the “Setting & props” sub-category, in particular the “Forest” group. This group includes 4 KWs: *bush, forest, hippo, and grass*. Among these, *forest* is immediately relevant: it occurs 23 times and it always refers directly to the aspect of the fictional world I aim to analyse. On the contrary, *bush, hippo, and grass* are less relevant. *Hippo* occurs only 5 times and refers mainly to the hippo-meat that a group

of natives bring with them on Marlow's steamboat – this is the reason why *hippo* also appears in the “Equipment & tools” sub-category. *Hippo* also refers once to an actual hippo living in the forest close to the station. *Bush* and *grass* occur more frequently (18 times each), but refer to partial aspects rather than to the African jungle as a whole. It can be said that *bush*, *grass*, and *hippo* are “props” of the setting “Forest”, rather than referring directly to it. Although they contribute to some extent to building up the fictional world, they do not reflect as specifically and comprehensively as *forest* the overall representation of the African forest.

Looking outside the “Places” sub-category, there are other three words that develop a similar function as *forest* and will therefore be included in the analysis. The first one is *jungle*. *Jungle* is used as a (near) synonym of *forest*, in the sense that it refers directly to the same physical place. However, *jungle* was not identified as a KW. This might be a consequence of the fact that *jungle* is not very frequent (it occurs only 6 times), considering that the outcome of a KW analysis depends on frequency figures. The *p*-value threshold used might have been too restrictive for the study of HoD. If the *p*-value cut-off point is increased from 0.0001 to 0.001, *jungle* becomes a KW when comparing HoD to CC, while increasing it to 0.01, it becomes a KW also when comparing HoD to CC. As a matter of fact, *jungle* is used over 7 times more frequently in HoD than in CC (normalised per 1,000 words) and almost 4 times more frequently than in RC. However, these differences are not statistically significant with a *p*-value as low as 0.0001. *Jungle* will equally be taken into consideration because, aside from its frequency, this word is as relevant as *forest* for the study of the theme “Africa and its representation”. *Forest* and *jungle* are used by Conrad with the exact same referent, and both describe the African forest in a concrete yet comprehensive way.

The remaining two words provide instead a less concrete and more metaphorical perspective on the representation of the African jungle. The first one is *wilderness*. I have argued in the previous section that *wilderness* is a polysemous word. In its concrete sense, this thematic signal KW is used to refer to the jungle, but at the same time it also has a more abstract meaning, i.e. the living soul of the African jungle that resists the Western intrusion. Including *wilderness* in the “Africa word” group enables me to develop the examination of the representation of the African

jungle on two levels: the concrete representation and the metaphorical representation. This is equally the case for *darkness*.

Although at first it might appear that *forest*, *jungle*, and *wilderness* have a different referent from *darkness* and that the former are not directly connected to the latter, *darkness* plays an important role in the representation of the African jungle. Its relation with the jungle derives from its multiple meanings, which are mirrored in the variety of different – and sometimes contrasting – ways the darkness in HoD has been interpreted. Similarly to *wilderness*, *darkness* has a concrete and straightforward interpretation which is directly related to the indomitable and wild heart of Africa. The very title of the short novel refers to the heart of darkness as the centre of the African continent, dark and mysterious. This link has been acknowledged by Goonetilleke (1999: 11) and O’Pray (1983: 12), who recognise the strong influence of Henry Morton Stanley’s accounts of his travels in Africa – *Through the Dark Continent* and *In the Darkest Africa* – on Conrad’s short novel. In line with this interpretation, *darkness* can be seen as directly relating with *forest* and *jungle*: although as a metaphorical expression, *darkness* refers to the same actual place alluded to by the other two words.

An alternative interpretation sees the darkness in HoD as an abstract notion rather than as a metaphorical reference to a physical place. In this sense, the darkness can convey numerous figurative meanings at the same time. Watts (1990) describes this darkness as “moral corruption, night, death, ignorance, and that encompassing obscurity of the pre-rational and pre-verbal which words seek to illuminate” (Watts 1990: xx-xxi), whereas O’Prey (1983) describes it as “the unknown, [...] the subconscious; it is also a moral darkness, it is the evil which swallows up Kurtz and it is the spiritual emptiness he sees at the centre of existence; but above all it is mystery itself, the mysteriousness of man’s spiritual life” (O’Prey 1983: 18). This interpretation sees the darkness linked to a multitude of either overt or covert themes, ranging from the colonial enterprise and its exploiting actions to the withering of the heart because of greed and selfishness resulting from a Mephistophelian exchange of soul for achievements. In this sense, there would be no direct link with Africa, therefore the connection of *darkness* with *forest* and *jungle* may be feebler, or even not existing. However, there are critics who put these two interpretations together. Goonetilleke (1999), for instance, interprets the darkness as both “the centre of

Africa” and, at the same time, “the unknown, the evil in humanity, the hidden self and the negation at the back of all things” (Goonetilleke 1999: 39).

As a thematic signal KW, *darkness* develops multiple meanings in HoD. Through this polysemy, it is possible to establish a connection between *darkness* and the African continent, as the idea of darkness is also used to refer metaphorically to the Congolese jungle. Including *darkness* in the analysis will allow me to investigate whether there are parallelisms with the use of *forest*, *jungle*, and *wilderness*, and will strengthen the double perspective on the metaphorical and concrete representations of the African jungle. Finally, the four items selected for the study of the fictional representation of the African jungle are: *forest*, *jungle*, *wilderness*, and *darkness*; henceforth referred to as the “Africa words”.

5.2 The “Africa words”

With respect to the fictional representation of the African natives, the selection of the items to analyse in detail starts from the “Characters” sub-category, specifically from the “Natives” group. In this case too, there are KWs that are not as relevant as others; thus, not all of the words in the group will be analysed. As for the representation of the African jungle, the aim of my analysis will be to examine the general depiction of the Africans, rather than just a partial aspect of their representation. Therefore, terms that refer to the natives as generally as possible are preferred over words that refer only to a selection of them. As a result, only 2 out of 5 words will be selected from the “Native” group: *savages* and *niggers*. Although used in different contexts, these terms refer to the natives in general, and not to specific individuals. By contrast, *cutters* and *helmsman* refer to precise characters. *Cutters* occurs 4 times, always in the compound noun *wood-cutters*, and refers to a small group of Africans that help Marlow in his journey through the wilderness. Similarly, *helmsman*, which occurs 7 times in total, refers to the African helmsman of Marlow’s steamer, a precise individual. *Chaps* is discarded because it is not used for the Africans specifically, but is rather a superordinate word adopted for other groups of characters as well, for example the colonisers (this is the reason why it also appears in the “Colonisers” group).

Besides *niggers* and *savages*, 7 other words will be selected in order to study the fictional representation of the Africans. These are: *nigger*, *negro*, *savage*, *native*, *natives*, *black*, *blacks*. There are three interrelated reasons for this further inclusion. The first reason parallels the decision to include *jungle* in the “Africa words”. *Nigger*, *negro*, *savages*, *native*, *black*, and their plural forms are all words Conrad uses to refer specifically to the African natives. Even though their frequency of use in HoD is not statistically significant with the chosen parameters, this does not mean that they are not relevant for this analysis. As I have hypothesised in reference to *jungle*, the choice of a different cut-off point or different selection criteria could have resulted in their inclusion among the KWs. *Nigger*, *savage*, and *black* are already KWs, but in List 2 only, while *savage* becomes a KW in List 1 if the *p*-value threshold is increased from 0.0001 to 0.001. Finally, adding *nigger*, *negro*, *savage*, *native*, *natives*, *black*, and *blacks* to the group ensures that all the general words employed in the text referring to the Africans are accounted for; in this way, the study of the representation of the natives is as comprehensive as possible.

The second reason for the inclusion of these terms is related to the first: the selection of a singular form leads to the selection of a plural form, and vice versa. Unlike the “Africa words”, both singulars and plurals are relevant in this case. While there is only one fictional entity to which the terms *forest*, *jungle*, *wilderness*, and *darkness* can refer, there are multiple natives, therefore it is important to consider *nigger*, *savage*, *native*, and *black* when they refer both to an individual and when they are used in the plural for a group of people (except for *negro*, which occurs only in its singular form). One could object that a word like *helmsman*, for example, has been discarded because it is used to identify an individual native, while here singular forms that equally refer to individuals are taken into account. The difference here is that, in the case of *helmsman*, the individual referred to is always the same character, whereas *nigger*, *savage*, or any other of these terms is used indistinctively for multiple characters as a general noun.

The third reason to support the inclusion of these additional words is that this methodological decision is in line with a study by Kujawska-Lis (2008), which has very similar aims to the present investigation. Kujawaska-Lis (2008) compares two Polish translations of HoD, one preceding and the other following Achebe’s (1990) public lecture. She investigates whether the contemporary translation differs from the

early one as a consequence of the influence that Achebe's (1990) interpretation had on the contemporary translator. In order to do so, Kujawaska-Lis (2008) focuses on the lexical choices used to translate the ST items referring to the African natives. In particular, she looks at the words *nigger*, *negro*, *black*, *savage*, *native*, *brute*, and their Polish translations. Kujawska-Lis (2008) shows that the lexical choices of the contemporary TT maintain or even emphasise the potential racist implications of the original. On the contrary, the early translator replaces the most derogatory terms with more neutral ones, tuning down the ST's potential racist implications. As a result, in the early translation, Marlow "is decidedly less verbally aggressive and mentally superior to the Africans" compared to the one portrayed in the contemporary TT, who is instead "biased both linguistically and intellectually" (Kujawska-Lis 2008: 176). She argues that, as a consequence of these alterations, the two TTs can have a different influence on the TT reader, for the early TT tunes down the possibility to read the text in line with Achebe's (1990) evaluation, while the contemporary one opens the door to it. Thus, Kujawska-Lis's (1990) study supports the choice of these terms as an appropriate selection to analyse the effect of translation on the fictional representation of the Africans.

Based on these three reasons, the 9 words selected to study the fictional representation of the African natives are: *nigger*, *niggers*, *negro*, *savage*, *savages*, *native*, *natives*, *black*, and *blacks*; henceforth referred to as the "African words".

6. Conclusion

This chapter has set the basis for the analysis in the following chapters, leading to the identification of relevant HoD-specific words that relate the lexical level of the text with the two major themes discussed in Chapter 2. As a result, this chapter has addressed Research Question 1. In the upcoming chapters, I will examine in detail the words within the two groups: in Chapter 5, I will focus on the representation of the African jungle through the examination of the "Africa words", while Chapter 6 will concentrate on the representation of the African natives through the investigation of the "African words".

The analysis of the words in these two groups allows me to account for both a text-based and a reader-based perspective on the major themes. On the one hand, some items (*forest, wilderness, darkness, savages, niggers*) are KWs in both lists; as such, they can be considered specific features of HoD from a text-based perspective. On the other hand, some other words (*jungle, nigger, savage, negro, native(s), and black(s)*) stand out because they are related to the specific themes I intend to study, and therefore reflect a reader-based understanding of the text (the shared and established reading of literary critics). The resulting groups, “Africa words” and “African words”, combine these two perspectives, providing a more nuanced understanding of how textual patterns relate to thematic concerns.

CHAPTER FIVE

“A place of darkness”: The fictional representation of the African jungle

1. Introduction

In this chapter, I will examine how the African jungle is linguistically represented in HoD and in its four Italian translations. The examination will be based on the outcome of the KW analysis and categorisation (Chapter 4) and it will focus in particular on the “Africa words” (*forest*, *jungle*, *wilderness*, and *darkness*). In Section 2, I will discuss the fictional representation of the African jungle as a “place of darkness”. I will start looking at *forest*, *wilderness*, and *jungle* as core words of functionally complete units of meaning (Section 2.1). I will show how these items enact together a local textual function which plays a key role in the building up of this specific aspect of the fictional world. In Section 2.2, I will concentrate on the analysis of *darkness*, showing how this word is strictly related to *forest*, *wilderness*, and *jungle*. I will argue that the four items jointly establish a cohesive image of Africa throughout the whole text. In Section 2.3, I will discuss this image in relation to the existing literary criticism, in order to contextualise the findings of the analysis and relate linguistic description and literary appreciation. I will argue that the lexico-semantic patterns shared by *forest*, *wilderness*, *jungle*, and *darkness* create a metaphorical representation of Africa as a “place of darkness”, a representation that is more figurative than concrete. This has important implications for the interpretation of the novel. Moreover, the analysis of the “Africa words” will also foreground a secondary tendency related to the representation of the jungle, that is, the depiction of the jungle as a personified entity. In Section 3, I will discuss this tendency, showing which linguistic features instantiate it and what stylistic effects they create.

From Section 4 on, the focus will be shifted to the TTs. In Section 4.1, I will compare the TTs and discuss how the alterations introduced in translation affect the representation of the jungle as a “place of darkness”. In Section 4.2, I will show that, although the lexico-semantic patterns identified in the original are maintained, some of the translators’ alterations have an effect on the lexical cohesion of the text. In Section 4.3, I will look instead at the effects of such alterations on the personification

of the jungle. In Section 4.4, I will consider the potential manipulative effects of such alterations. Taking into account the usual understanding and definition of ‘manipulation’ in translation studies research, in Section 4.4 I will discuss whether the alterations identified can be regarded as an example of manipulation or not. I will argue that manipulation can manifest itself also through obligatory linguistic shifts, which are usually not considered instances of manipulation. Finally, in Section 5 I will provide some concluding remarks.

2. The African jungle as a “place of darkness”

In this chapter, I will discuss two tendencies in the linguistic construction of Africa in HoD. The main tendency involves representing the jungle as a dark, immense, and impenetrable place: this representation will be labelled “place of darkness”. In this section, I will study this representation, showing the stylistic patterns that instantiate it and discussing its interpretational implications.

2.1 *Forest, jungle, and wilderness*

Forest, jungle, and wilderness occur respectively 23, 6, and 22 times. They are all used to refer to the same aspect of the fictional world, namely the African jungle, although in different ways. Whereas *forest* and *jungle* label the place concretely, *wilderness* is used both concretely and metaphorically, as I have argued in Sections 4 and 5 of the previous chapter. However, the analysis of their concordance lines shows that the three words are functionally related, as all of them co-occur with the same semantic fields, that is, they share the same semantic preferences. Consider the concordance lines of *forest* in Figure 5.1 below.

1	that mysterious life of the wilderness that stirs in the forest, in the jungles, in the hearts of wild men. There's no
2	shoulder-blades. Then the whole population cleared into the forest, expecting all kinds of calamities to happen, while,
3	his short flipper of an arm for a gesture that took in the forest, the creek, the mud, the river--seemed to beckon
4	fitting indistinctly against the gloomy border of the forest, and near the river two bronze figures, leaning on tall
5	took these sticks to bed with them. Beyond the fence the forest stood up spectrally in the moonlight, and through that
6	that I leaped to my feet and looked back at the edge of the forest, as though I had expected an answer of some sort to
7	The reaches opened before us and closed behind, as if the forest had stepped leisurely across the water to bar the way
8	deck. A frightful clatter came out of that hulk, and the virgin forest on the other bank of the creek sent it back in a
9	stream. It was then well on in the afternoon, the face of the forest was gloomy, and a broad strip of shadow had already
10	there had been between, had disappeared. Of course the forest surrounded all that. The river-bank was clear, and on
11	persistently with his whole arm. Examining the edge of the forest above and below, I was almost certain I could see
12	without any perceptible movement of retreat, as if the forest that had ejected these beings so suddenly had drawn
13	by Jove! was in my nostrils, the high stillness of primeval forest was before my eyes; there were shiny patches on the
14	itself in the calm of the evening. The long shadows of the forest had slipped downhill while we talked, had gone far
15	were poured into the clearing by the dark-faced and pensive forest. The bushes shook, the grass swayed for a time,
16	Station. It was on a back water surrounded by scrub and forest, with a pretty border of smelly mud on one side, and
17	kings. An empty stream, a great silence, an impenetrable forest. The air was warm, thick, heavy, sluggish. There was
18	because on a certain occasion, when encamped in the forest, they had talked all night, or more probably Kurtz had
19	I was convinced, had driven him out to the edge of the forest, to the bush, towards the gleam of fires, the throb of
20	was keeping guard over the ivory; but deep within the forest, red gleams that wavered, that seemed to sink and
21	the trees, and the murmur of many voices issued from the forest. I had cut him off cleverly; but when actually
22	but as a rule Kurtz wandered alone, far in the depths of the forest. Very often coming to this station, I had to wait days
23	and was looking at the shore, sweeping the limit of the forest at each side and at the back of the house. The

Figure 5.1 Concordance lines of *forest*

Forest is firstly described as a place closed in its boundaries, impenetrable from the outside or inescapable once inside. This is conveyed by words such as *fence* (line 5), *closed* (line 7), *surrounded* (lines 10 and 16), and *impenetrable* (line 17), together with the construction *the * of the forest*, where the wildcard position is occupied by terms for borders and edges: *border* (line 4), *edge* (lines 6, 11, and 19), and *limit* (line 23). I will call this semantic field “impenetrableness”. Example (1) shows the entire sentence in line 7, to demonstrate how this aspect of the representation of the African forest is enacted in a longer extract. This passage not only emphasises that the forest has borders, but also that once these borders have been crossed, it is difficult to get out:

- (1) The reaches opened before us and closed behind, as if the forest had stepped leisurely across the water to bar the way for our return. We penetrated deeper and deeper into the heart of darkness.

The second dominant aspect in the description of *forest* is represented by the semantic field “darkness”, conveying the idea of the forest as a dark place, both literally and figuratively. This field includes words such as *gloomy* (lines 4 and 9), *spectrally* (line 5), *shadows* (line 14), and *dark-faced* (line 15), but also *primeval* (line 13) and *silence* (line 17) depict the forest as an altogether inhospitable place. In Example (2) below, the contrast between the darkness of the forest and the faint light of the colonisers’ station is made explicit. The shadows of the forest are slowly surrounding the station, the only place where the sunshine still resists:

- (2) The long shadows of the forest had slipped downhill while we talked, had gone far beyond the ruined hovel, beyond the symbolic row of stakes. All this was in the gloom, while we down there were yet in the sunshine, and the stretch of the river abreast of the clearing glittered in a still and dazzling splendour, with a murky and overshadowed bend above and below.

The third semantic field is related to the other two, as it combines with the idea of “impenetrableness” and “darkness” to depict the forest as an immense place, almost shapeless, in which it is very easy to get lost. I will call this semantic field “immensity”. In the case of *forest*, it is instantiated by the words *deep* (line 20) and *depths* (line 22). It becomes more evident if we look at the concordance lines of *wilderness*. See Figure 5.2 below.

1	home--perhaps; setting his face towards the depths of the wilderness , towards his empty and desolate station. I did
2	had closed round him--all that mysterious life of the wilderness that stirs in the forest, in the jungles, in the
3	one would think, they were lugging, after a raid, into the wilderness for equitable division. It was an inextricable
4	well equipped for a renewed encounter with the wilderness . 'Ah! I'll never, never meet such a man again.
5	I tried to break the spell--the heavy, mute spell of the wilderness--that seemed to draw him to its pitiless breast
6	himself, afterwards he arose and went out--and the wilderness without a sound took him into its bosom again.
7	little touch of other things--the playful paw-strokes of the wilderness , the preliminary trifling before the more serious
8	but this--ah--specimen, was impressively bald. The wilderness had patted him on the head, and, behold, it
9	kept him unscathed. He surely wanted nothing from the wilderness but space to breathe in and to push on
10	which went rotten, and made the mystery of the wilderness stink in my nostrils. Phoo! I can sniff it now. I
11	It made me hold my breath in expectation of hearing the wilderness burst into a prodigious peal of laughter that
12	but I will wring your heart yet! he cried at the invisible wilderness . "We broke down--as I had expected--and had
13	a few days the Eldorado Expedition went into the patient wilderness , that closed upon it as the sea closes over a
14	suddenly upon the whole sorrowful land, the immense wilderness , the colossal body of the fecund and
15	no going ashore. Here and there a military camp lost in a wilderness , like a needle in a bundle of hay--cold, fog,
16	clerks to levy toll in what looked like a God-forsaken wilderness , with a tin shed and a flag-pole lost in it;
17	anything so unreal in my life. And outside, the silent wilderness surrounding this cleared speck on the earth
18	noise. But his soul was mad. Being alone in the wilderness , it had looked within itself, and, by heavens! I
19	conquering darkness. It was a moment of triumph for the wilderness , an invading and vengeful rush which, it
20	She stood looking at us without a stir, and like the wilderness itself, with an air of brooding over an
21	have at least a choice of nightmares. "I had turned to the wilderness really, not to Mr. Kurtz, who, I was ready to
22	came to him at last--only at the very last. But the wilderness had found him out early, and had taken on

Figure 5.2 Concordance lines of *wilderness*

In Figure 5.2, the “immensity” semantic field is clearer. *Wilderness* too occurs with *depths* (line 1); it is also an *immense* and *colossal* place (line 14), in which camps and stations are *lost* (line 15). In line with the “impenetrableness” semantic field identified in the concordance lines of *forest*, *wilderness* is described as a sentient being that surrounds and closes in whoever dares to challenge its impenetrableness (*closed*, line 13; *surrounding*, line 17). I will return and expand on the agency of *wilderness* in Section 2.3. As with *forest*, *wilderness* is also described as an inhospitable and dark place. The “darkness” semantic field is identifiable in the co-occurrence of words such as *desolate* (line 1), *mysterious* (line 2), *mystery* (line 10), *sorrowful* (line 14), *God-forsaken* (line 16), *vengeful* (line 19), and *brooding* (line 20).

Even though *jungle* occurs only 6 times in the text, it is equally possible to recognise a connection with *forest* and *wilderness*, for *jungle* too is described in terms of the three semantic fields identified. Consider Figure 5.3 below. As for the “impenetrableness” semantic field, the jungle is represented as an *impenetrable* (line 1) place closed in its *edge* (line 4); as for the “immensity” semantic field, the jungle is

also *colossal* (line 4) and *immense* (line 5); finally, as for the “darkness” semantic field, the jungle is described as *dark* and *black* (line 4).

1	be if we attempted to move. Still, I had also judged the jungle of both banks quite impenetrable—and yet eyes
2	his talk of chains and purchases, made me forget the jungle and the pilgrims in a delicious sensation of having
3	you that never, never before, did this land, this river, this jungle , the very arch of this blazing sky, appear to me so
4	aspect of monotonous grimness. The edge of a colossal jungle , so dark-green as to be almost black, fringed with
5	of the towering multitude of trees, of the immense matted jungle , with the blazing little ball of the sun hanging over
6	large holes in the peaked roof gaped black from afar; the jungle and the woods made a background. There was no

Figure 5.3 Concordance lines of *jungle*

These three semantic fields are not simply shared by and repeated for *forest*, *wilderness*, and *jungle*. The pervasiveness of this specific fictional representation is further confirmed by the fact that the semantic preferences are also realised by forms of the same lemmas or even identical words. *Depths* occurs with *forest* and *wilderness*; both occur with forms of the lemma SURROUND (*surrounded* and *surrounding*); *impenetrable* occurs with *jungle* and *forest*. Both *wilderness* and *jungle* are *immense* and *colossal*, while *jungle* and *forest* are *dark*. This reiteration of lexical choices in relation to Africa strengthens the connection of the three words and suggests that they participate together in the construction of a unified fictional entity. For ease of reference, I will adopt the label “place of darkness” to define the fictional representation of the African jungle resulting from the three semantic preferences, even though this name does not cover all the nuances expressed by the individual semantic fields (“darkness”, “impenetrableness”, and “immensity”). The name for the label “place of darkness” is suggested by Marlow himself: at the beginning of the book, he recognises that Africa, once a blank space on the world map, still unexplored, now is instead “a place of darkness” (Conrad [1902] 1994: 12).

As further evidence to the specificity to this fictional representation, I will check whether *forest*, *wilderness*, and *jungle* co-occur with the same words discussed above in the two reference corpora, taking into account a similar span of 5:5. This test is not meant to prove that these co-occurrences represent statistically significant collocations, or that their collocational strength is stronger in HoD than in the two reference corpora. As I have explained in Chapter 3, the analysis in this and in the following chapter does not focus on the quantitative aspects of the lexical item. I have argued (Chapter 3, Section 4.2.1) that a strictly quantitative approach would cut out

linguistic features that, despite their low frequency of occurrence, contribute distinctively to the construction of the two major themes under analysis. In this specific case, the words that form part of the three semantic fields identified would hardly be considered statistical collocations of *forest*, *wilderness*, and *jungle*, because they occur only a few times in the whole text. However, there is no frequency threshold for a word to be part of a semantic field, therefore semantic preferences and prosodies allow me to account for the cumulative effect of items whose frequency is not necessarily high.

For example, if mutual information (Baker 2010: 24-25) were used as a statistical test to calculate the collocations of *forest*, *wilderness*, and *jungle*, it would have been difficult to identify the semantic fields. Mutual information (MI) is one of the most used statistical tests to calculate the collocational strength between two words. Because it tends to assign high MI scores to words with relatively low frequency (Baker 2006: 102), a frequency cut-off point is generally used to reduce this drawback (Xiao & McEnery 2006: 105). Even a frequency cut-off point as low as 3 would have excluded most of the words that form part of the “darkness”, “impenetrableness”, and “immensity” semantic fields, if MI was adopted to retrieve the collocations of *forest*, *wilderness*, and *jungle*.

Therefore, the terms ‘collocation’ and ‘collocate’ are used here in a non-statistical sense or, in McEnery & Hardie’s (2012: 126) words, as ‘collocation-via-concordance’, that is, collocations identified by eye through the analysis of concordance lines without statistical significance testing. Checking whether *forest*, *wilderness*, and *jungle* occur with the same words in CC and RC, although without statistical support, is equally relevant, as it can shed light on how text-specific this representation of the jungle is. Consider Tables 5.1 to 5.3 below, where a grey cell is used to indicate when the co-occurrence of *forest*, *wilderness*, and *jungle* with a word from the three semantic fields does not appear in the relative reference corpus.

Table 5.1 Comparison of the collocates of *forest*

Collocates of <i>forest</i>	CC	RC
border	No	No
closed	Yes	No
dark	Yes	Yes
deep	Yes	Yes
depths	No	No
edge	Yes	Yes
fence	No	No
gloomy	Yes	No
impenetrable	Yes	No
limit	Yes	No
primeval	Yes	No
shadows	Yes	Yes
side	Yes	Yes
silence	Yes	Yes
spectrally	No	No
surrounded	No	No

Table 5.2 Comparison of the collocates of *jungle*

Collocates of <i>jungle</i>	CC	RC
colossal	No	No
impenetrable	No	No
edge	No	No
immense	No	No
dark	Yes	No
black	No	Yes

Table 5.3 Comparison of the collocates of *wilderness*

Collocates of <i>wilderness</i>	CC	RC
surrounding	Yes	No
God-forsaken	No	No
invisible	No	No
colossal	No	No
sorrowful	No	No
mysterious	No	No
closed	No	No
depths	No	No
immense	Yes	No
mystery	No	No
lost	Yes	No

The three tables show that, in most cases, *forest*, *wilderness*, and *jungle* do not occur with the words creating the “place of darkness” representation in the reference

corpora. The only exception is in Table 5.1, where the majority of the collocates appear in a 5:5 span from *forest* in CC. This indicates that Conrad uses the same words to describe forests also in his other works. However, apart from that, Tables 5.1 to 5.3 suggest that the “impenetrableness”, “darkness”, and “immensity” semantic preferences are text-specific, that they are more typical of the fictional representation of the African jungle in HoD than of Conrad’s fictional language or the general literary language of the period.

2.2 The fictional representation of *darkness*

The fourth item of the “Africa words” is *darkness* (cf. Section 5.1 of Chapter 4 for the selection rationale). I have argued that, similarly to *wilderness*, *darkness* is a polysemous thematic signal KW that carries both a figurative and a concrete meaning. It can either connote “the unknown, the evil in humanity, the hidden self and the negation at the back of all things” (Goonetilleke 1999: 39) or simply refer to the heart of the African jungle.

This polysemous nature is foregrounded by Turci’s (2007) analysis of *dark** (*dark*, *darkness*, *darker*, and *darkly*) in HoD. Turci (2007) links quantitative corpus methods to qualitative systemic functional linguistic analysis to identify six categories which *dark** refers to: “European people”, “African people”, “Kurtz and/or Kurtz character merging with African landscape”, “Kurtz’s African mistress”, “European places”, and “Africa and the Congo”. Turci (2007: 109) shows that, among these, *dark** refers most commonly to Kurtz, to European places, and to Africa and the Congo, confirming that *dark** can equally relate to different aspects of the short novel.

The direct connection with the African landscape that Turci (2007) points out is further confirmed by my analysis. *Darkness* in fact shares the very same “place of darkness” representation as the other “Africa words”. Consider Figure 5.4 below, which shows all of 25 concordance lines of *darkness*.

1	arms over the glitter of the infernal stream, the stream of darkness . She said suddenly very low, 'He died as he
2	to the lurking death, to the hidden evil, to the profound darkness of its heart. It was so startling that I leaped to
3	mind with its hint of danger that seemed, in the starred darkness , real enough to make me get up for the purpose
4	even to know you are being assaulted by the powers of darkness . I take it, no fool ever made a bargain for his
5	was to know what he belonged to, how many powers of darkness claimed him for their own. That was the
6	return. We penetrated deeper and deeper into the heart of darkness . It was very quiet there. At night sometimes the
7	died I have had no one--no one--to--to--' 'I listened. The darkness deepened. I was not even sure whether he had
8	enough to penetrate all the hearts that beat in the darkness . He had summed up--he had judged. The
9	with an unearthly glow in the darkness, in the triumphant darkness from which I could not have defended her--from
10	saving illusion that shone with an unearthly glow in the darkness , in the triumphant darkness from which I could
11	in his time, perhaps. They were men enough to face the darkness . And perhaps he was cheered by keeping his
12	earth, the unseen presence of victorious corruption, the darkness of an impenetrable night.... The Russian tapped
13	far away there I thought of these two, guarding the door of Darkness , knitting black wool as for a warm pall, one
14	it last as long as the old earth keeps rolling! But darkness was here yesterday. Imagine the feelings of a
15	like the beating of a heart--the heart of a conquering darkness . It was a moment of triumph for the wilderness,
16	a voice speaking from beyond the threshold of an eternal darkness . 'But you have heard him! You know!' she cried.
17	going at it blind--as is very proper for those who tackle a darkness . The conquest of the earth, which mostly
18	smile of his, as though it had been a door opening into a darkness he had in his keeping. You fancied you had
19	to hide in the magnificent folds of eloquence the barren darkness of his heart. Oh, he struggled! he struggled! The
20	cottons, beads, and brass-wire set into the depths of darkness , and in return came a precious trickle of ivory.
21	smoke. "The brown current ran swiftly out of the heart of darkness , bearing us down towards the sea with twice
22	a boy to dream gloriously over. It had become a place of darkness . But there was in it one river especially, a
23	sky--seemed to lead into the heart of an immense darkness . arkness.
24	or the deceitful flow from the heart of an impenetrable darkness . "The other shoe went flying unto the devil-god
25	of my ideas. It's a duty." "His was an impenetrable darkness . I looked at him as you peer down at a man

Figure 5.4 Concordance lines of *darkness*

Darkness too co-occurs with the “impenetrableness” semantic field. The darkness is *impenetrable* (lines 12, 25, and 25) and described in terms of something you go in or out of: *door* (lines 13 and 18), *threshold* (line 16), and *opening* (line 18). In line 22, Africa is directly referred to as *a place of darkness*. The concordance lines also show the presence of the “immensity” semantic field. *Darkness* co-occurs with words such as *profound* (line 2) and *immense* (line 23), but also *deeper* (line 6), *deepened* (line 7), and *depths* (line 20). Finally, it is also possible to identify the “darkness” semantic field, as conveyed by items such as *infernal* (line 1), *evil* (line 2), *assaulted* (line 4), *unearthly* (line 10), and *corruption* (line 12). What is more, the three semantic fields include again many of the exact same words and lemma forms used with *wilderness*, *jungle*, and *forest*. *Darkness*, *forest*, and *jungle* are *impenetrable* (among the most occurring collocates); *jungle*, *wilderness*, and *darkness* are *immense*; forms of the lemma DEEP co-occur with *darkness* (*deeper*, *depths*, *deepened*), *forest* (*deep*, *depths*), and *wilderness* (*depths*).

Differently from *forest*, *jungle*, and *wilderness*, the majority of the collocates of *darkness* do not seem to be typical of HoD (see Table 5.4). *Darkness* co-occurs with the same words also in CC and in RC, suggesting that relations such as *darkness-impenetrable*, *darkness-depths*, and *darkness-profound* perhaps are not text-specific.

Table 5.4 Comparison of the collocates of *darkness*

Collocates of <i>darkness</i>	CC	RC
deepened	Yes	Yes
impenetrable	Yes	Yes
deeper	Yes	Yes
threshold	Yes	No
infernal	No	No
evil	No	Yes
unearthly	No	No
opening	No	Yes
door	Yes	Yes
depths	Yes	Yes
profound	Yes	Yes
immense	Yes	Yes
assaulted	No	No
corruption	No	No

However, more than the specificity of these collocations, what is relevant for the present study is the fact that *darkness* is described in similar terms as *forest*, *jungle*, and *wilderness*. My analysis shows that a link exists between *darkness* on the one hand and *forest*, *wilderness*, and *jungle* on the other. This link strengthens the relation between the metaphorical understanding of *darkness* and the actual depiction of Africa.

Overall, all four “Africa words” show the same lexico-semantic patterns. These patterns convey a fundamental aspect of the fictional representation of the African jungle. Every time *forest*, *wilderness*, *jungle*, and *darkness* occur, the reader encounters the same “place of darkness” fictional representation, enacted through the use of the same words, lemma forms, and semantic fields. The repetition of semantically related vocabulary to refer to the same aspect of the fictional world throughout the text makes the depiction of the African jungle cohesive. A network is established at both the lexical and semantic level, connecting the occurrences of the four items. Given that lexical cohesion (Halliday & Hasan 1976: 274) is based upon

the choice of lexis in organising relations within texts, the use of the same words and semantic fields for each occurrence of the four “Africa words” builds up a unified and consistent fictional entity: “the use of a lexical item recalls [through lexical chains] the sense of an earlier one” (Baker 2011: 211), creating in this way a thick network of lexical cohesion.

2.3 Africa as a metaphorical place

In the previous sections, I have argued that the use of *forest*, *wilderness* and *jungle* on the one hand, and *darkness* on the other, are interrelated. Their shared textual behaviour enacts a unified and cohesive representation of the African jungle in which the depiction of the physical place is linked to the figurative meaning of *darkness* and *wilderness*. These linguistic features construct the African jungle as a “place of darkness”. It is dark and inhospitable (the “darkness” semantic field, e.g. *dark*, *gloomy*, *shadows*, *black*, *desolate*, *mysterious*, *mystery*, *sorrowful*, *silent*, *spectrally*), immense and with blurred boundaries (the “immensity” semantic field, e.g. *immense*, *colossal*, *deep*, *depths*, *lost*) but at the same time it still resists any attempt of intrusion from the outside (the “impenetrableness” semantic field, e.g. *edge*, *border*, *fence*, *impenetrable*, *surrounded*, *closed*). In the end, what these consistent lexico-semantic patterns enact is a symbolic and metaphorical representation of the African jungle, rather than a geographical or descriptive one.

I have explained in Chapter 4 (Section 4) that Conrad’s writing has been considered vague and indefinite. Leavis (1973: 180) has famously argued that Conrad “mak[es] a virtue out of not knowing what he means”, while Stubbs (2005: 8) has pointed out that HoD is full of “examples of this vagueness about places, times, and people”. As far as places are concerned, the KW categorisation has shown that common place nouns are much more frequent than proper place names. As a matter of fact, Congo is never named in the book, while the noun *Africa* appears only once, in the passage where Marlow refers to it as “a place of darkness”. Apart from that, there is no other clearly defined geographical reference to the country. This lack of a concrete geographical focus is in agreement with what the analysis of the fictional representation of the African jungle has shown so far. The way in which the jungle is

referred to shifts the depiction of the continent from a physical description to a figurative representation, contributing further to the overall feeling of vagueness.⁷ Challenging any attempt of spatialisation and geographical contextualisation, Conrad delivers a fictional representation of the jungle that turns Africa into a metaphorical space, into a symbol of indefiniteness and blankness.

Brooker & Thacker (2005: 4-5) recognise this practice of viewing material places only through the lens of aesthetic metaphors as typical of literary modernism. In line with this, many other critics have identified a shift from a geographical into a symbolic space in the construction of Africa in HoD. Hegglund (2005: 43) argues that Africa in Conrad is transmuted from referent to signifier, from content to form, in a way that separates the idea of Africa from its material reality. The result of this process is a homogenising abstraction that deprives the continent of its cultural-historical variety. This shapeless abstraction, according to Hegglund (2005), is emphasised through contrast, when it is used as a term of comparison to set against any number of (Western) symbols and values (Hegglund 2005: 48). In a similar way, Achebe (1990: 3) states that Africa in HoD ceases to be an actual place in order to become a symbolic place, “the other world”, the antithesis of Europe. In this symbolic form, functioning almost just as a setting and backdrop, it loses its human and cultural factors. Goonetilleke (1999: 42-43) also points out that, rather than being the practical description of an expedition, the journey into Congo becomes a symbolic journey into the darkness of man’s unconsciousness, whereas Levenson (2009: 185) underlines that, as many other modernist writers, Conrad sets aside the wider context of the living culture, selecting instead only some aesthetic images of Africa. Finally, Hampson (2005: 56) compares HoD with Conrad’s notes and diary from his journey to Congo to emphasise how the Africa depicted in HoD resists all attempts of mapping its space.

⁷ Some readers might be interested to learn that this figurative rather than concrete representation of the African jungle was not the result of a lack of knowledge of Africa on the part of Conrad. Decades of expertise as professional seafarer and his personal experience in the very place he describes in HoD gave him first-hand knowledge of both Congo and how to geographically situate this place. Hampson (2005) suggests that Conrad was surely at ease with maps and mapping, with geographical and spatial coordinates. This is further confirmed by Conrad’s *Congo Diary*, in which his observations of the landscape and of the route followed during the journey reveal unquestionably his trained eye and professional skills (Hampson 2005: 57). Therefore, Conrad seems to have chosen not to represent the African jungle on a factual basis, even though he had the knowledge and skills to do so. Another example of this tendency is the fact that Conrad exaggerates, historically speaking, the primitiveness and isolation of his fictional Congo by predating it (O’Pray 1983: 15). In reality, the country was much less underdeveloped in 1890 than the way Conrad portrays it (Goonetilleke 1999: 10).

The textual patterns that I have identified enact exactly this local textual function: they construct a fictional representation of the African jungle that is more figurative than descriptive. The shared lexico-semantic features across all four “Africa words” make the representation of Africa cohesive and persistent throughout the text. Finally, Africa ceases to be a geographical place and becomes instead a blank, empty space, a “place of darkness”: the stylistic features identified are the lexical tools through which this representation is instantiated.

3. The fictional representation of the African jungle as a personified entity

The analysis of the “Africa words” has pointed out an additional aspect of the fictional representation of the African jungle, besides its depiction as a “place of darkness”: the personification of the forest, attributing to it agency and cognitive abilities. ‘Personification’ (or, sometimes, ‘prosopopoeia’) is defined as a figure of speech involving the attribution of human features to inanimate objects, animate non-human beings, or abstract entities (see Wales 2011 or Mortara Garavelli 2010). It is an “ontological metaphor” (Dancygier & Sweetser 2014: 62), as it alters the ontological status of an entity, often shifting it from inanimate to animate. Personification is commonly used in literary language to add vividness to a rhetorical discourse (Wales 2011: 314) and to represent figuratively the way an inanimate or abstract entity affects the speaker (Dancygier & Sweetser 2014: 63). Instances of personification have been already recognised in Conrad’s work and in HoD in particular. Miller (2006), for example, identifies prosopopoeia as one of most ostentatious literary features of HoD. Present even in the title, where a heart is ascribed to the darkness, personification is a key element in “naming by indirection what Conrad calls, in a misleading and inadequate metaphor, ‘the darkness’ or ‘the wilderness’” (Miller 2006: 467).

Out of the four “Africa words”, it is *forest* and *wilderness* that markedly show this tendency, identifiable in many of their concordance lines (see Figure 5.5 below).

1	in their hands. I verily believe they took these sticks to bed with them. Beyond the fence the forest stood up spectrally in the moonlight, and through that dim stir, through the faint so
2	of its heart. It was so startling that I leaped to my feet and looked back at the edge of the forest, as though I had expected an answer of some sort to that black display of confider
3	leaking we crawled very slow. The reaches opened before us and closed behind, as if the forest had stepped leisurely across the water to bar the way for our return. We penetrate
4	a jig. We capered on the iron deck. A frightful datter came out of that hulk, and the virgin forest on the other bank of the creek sent it back in a thundering roll upon the sleeping st
5	tree projected rigidly over the stream. It was then well on in the afternoon, the face of the forest was gloomy, and a broad strip of shadow had already fallen on the water. In this s
6	carved balls. The rails, or whatever there had been between, had disappeared. Of course the forest surrounded all that. The river-bank was clear, and on the waterside I saw a white r
7	that the crowd of savages was vanishing without any perceptible movement of retreat, as if the forest that had ejected these beings so suddenly had drawn them in again as the breath i
8	glances and savage movements, were poured into the clearing by the dark-faced and pensive forest. The bushes shook, the grass swayed for a time, and then everything stood still in
1	little use for any practical purpose. I tried to break the spell--the heavy, mute spell of the wilderness--that seemed to draw him to its pitiless breast by the awakening of forgotten and
2	looking very sick and trying to recover himself; afterwards he arose and went out--and the wilderness without a sound took him into its bosom again. As I approached the glow from the
3	. I had often 'a little fever,' or a little touch of other things--the playful paw-strokes of the wilderness, the preliminary trifling before the more serious onslaught which came in due course
4	say the hair goes on growing sometimes, but this--ah--specimen, was impressively bald. The wilderness had patted him on the head, and, behold, it was like a ball--an ivory ball; it had
5	, my--' everything belonged to him. It made me hold my breath in expectation of hearing the wilderness burst into a prodigious peal of laughter that would shake the fixed stars in their
6	.' I did so. There was a silence. 'Oh, but I will wring your heart yet!' he cried at the invisible wilderness. 'We broke down--as I had expected--and had to lie up for repairs at the head of
7	without bending a single blade. 'In a few days the Eldorado Expedition went into the patient wilderness, that closed upon it as the sea closes over a diver. Long afterwards the news
8	. And in the hush that had fallen suddenly upon the whole sorrowful land, the immense wilderness, the colossal body of the fecund and mysterious life seemed to look at her, pensive
9	some corpse. By Jove! I've never seen anything so unreal in my life. And outside, the silent wilderness surrounding this cleared speck on the earth struck me as something great and
10	beating of a heart--the heart of a conquering darkness. It was a moment of triumph for the wilderness, an invading and vengeful rush which, it seemed to me, I would have to keep back
11	himself I can't say, I think the knowledge came to him at last--only at the very last. But the wilderness had found him out early, and had taken on him a terrible vengeance for the

Figure 5.5 Concordance lines of *forest* and *wilderness*: personification

At times, *forest* and *wilderness* are depicted as animate entities able to move physically, as opposed to inanimate and immovable places. Their movements are actively produced, and not just the passive consequence of an external factor. *Forest* and *wilderness* are indeed given active agency through functioning as actor of material clauses. For example, the forest *stood up* in line 1, or *stepped across* in line 3. Similarly, the wilderness in line 4 *had patted* Kurtz on his head, while in line 7 it *closed upon* the Eldorado Expedition. Other times, *wilderness* and *forest* seem to possess their own will and cognitive skills. In line 3 the forest *leisurely* steps across the steamer to *bar the way for our return*. Here an intention is recognisable behind the action of the forest, which acts as moved by leisure. This “thoughtful” attitude is distinguishable also in the use of the adjective *pensive*, used with both *forest* (line 8) and *wilderness* (line 8) that, as physical spaces, are not expected to have the ability to think. However, both Marlow and Kurtz seem to have this expectation, as they verbally interact with the African jungle. In line 2, Marlow expects the forest to give him some sort of *answer* for his actions, whereas in line 6 Kurtz speaks out loud to the *invisible wilderness*. At times the wilderness seems even to be on the verge of interacting back, as in line 5, where the wilderness *burst into a prodigious peal of laughter*. Laughing can be seen as a behavioural process that typically pertains to humans (Halliday & Matthiessen 2004: 248). Its use in relation to the wilderness contributes to confer to it human traits. This is equally true for those cases in which

forest and *wilderness* are involved in mental processes. In a clause of mental process, “there is always one participant who is human” or, more precisely, a “being ‘endowed with consciousness’” (Halliday & Matthiessen 2004: 201). Yet the entity endowed with consciousness in these examples is not human, but rather inanimate, i.e. the wilderness. In line 8, the wilderness seems *to look at* Kurtz’s assumed mistress, in a *pensive* way; in line 11, it *found out* Kurtz’s *fantastic invasion* and takes vengeance on him.

The personifying tendency is instantiated not only through assigning human actions and behaviours to the African jungle, as seen so far, but also through the attribution of descriptive features and traits that are more animal- or human-like than specific to inanimate entities. An example of an animal attribute assigned to the wilderness is recognisable in line 3, where *wilderness* appears to possess paws. However, more pervasive than the beastly representation, is the depiction of the African forest as a woman. Apart from the ungendered *face* that co-occurs twice with *forest*, in lines 5 and 8, the forest is described as a *virgin* in line 4. Additionally, a *breast* is attributed to the wilderness in line 1, while in line 2 the wilderness takes Kurtz into its *bosom*. Finally, this depiction is strengthened up also through the use of *fecund* in reference with the colossal body of the wilderness, in line 8.

Overall, both actions and descriptions of the jungle as an animate and conscious entity ascribe hostile intentions to its personified representation, rather than benign ones. When the forest stands up, it does it *spectrally* (line 1); its face is *gloomy* (line 5) and *dark* (line 8). Its actions seem to be directed at interfering with Marlow’s expedition or, more generally, with all human intrusions altogether. As I have underlined before, the forest *leisurely* tries to *bar the way* of the steamer (line 3), and both *forest* and *wilderness* co-occur with forms of the lemma SURROUND (*surrounded* in line 6 of *forest*; *surrounding* in line 9 of *wilderness*), which conveys the idea of an intentional effort to isolate intruders. The wilderness seems to laugh at Kurtz (line 5) when he dares to challenge its dominion, and then takes a *terrible vengeance* (line 11, but also *vengeful* in line 10) on him for his imprudence. Generally speaking, this matches the “place of darkness” fictional representation, especially in its most obscure traits (the “darkness” and “impenetrableness” semantic preferences).

Now that the main trends of the personifying tendency have been identified in the context of *wilderness* and *forest*, I will analyse some textual extracts, so as to

show how the personification trope is developed in longer passages. Consider Example (3) below:

- (3) The wilderness had patted him on the head, and, behold, it was like a ball – an ivory ball; it had caressed him, and – lo! – he had withered; it had taken him, loved him, embraced him, got into his veins, consumed his flesh, and sealed his soul to its own by the inconceivable ceremonies of some devilish initiation.

In this extract, which provides extended context for line 4 of the concordance lines of *wilderness* (Figure 5.4), the metaphor of the wilderness as a woman is developed further. The image is that of a seductive yet dangerous partner who bewitches Kurtz, luring him into its *devilish initiation*. The lexical choices depict an almost physical encounter between lovers, whereby a personified wilderness caresses, embraces, and finally loves Kurtz, although at the same time it consumes his flesh and withers his scalp, sealing his soul to its own. This relation between the wilderness and its feminine embodiment is made most explicit when Marlow meets Kurtz's African mistress. Consider the extract in Example (4) below:

- (4) And in the hush that had fallen suddenly upon the whole sorrowful land, the immense wilderness, the colossal body of the fecund and mysterious life seemed to look at her, pensive, as though it had been looking at the image of its own tenebrous and passionate soul.

From this passage (the extended context for line 8 of the concordance lines of *wilderness* in Figure 4.5) it is clear that not only the wilderness has *its own tenebrous and passionate soul*, but that this soul finds its physical embodiment in an African woman. In fact, immediately before these lines, the African woman is described as *savage and superb, wild-eyed and magnificent; there was something ominous and stately in her deliberate progress*, a description that would work similarly well for the African jungle as it is fictionally represented in HoD.

4. The fictional representation of the African jungle in translation

In the previous sections, I have shown how the “Africa words” develop recurrent lexico-semantic patterns which build up the “where” of HoD. The lexical and semantic features identified transmute the real and geographical identity of Africa into a figurative representation. This representation reflects the socio-cultural context it originates from: it is hardly a neutral depiction, but rather mirrors historical values and cultural prejudices. Therefore, these patterns have a double value: first, they are an integral part of the style of the text; second, they possess critical relevance, because they play a key role in the interpretation of the novel. In this section, I will examine whether and how these patterns have been reproduced in translation. I argue that their alteration would affect both the style of the text and its interpretation.

I will start comparing the TTs, in order to study whether an alternation of such patterns have taken place in translation or not. Building on the findings of the analysis of the ST, this section discusses the effects of translation on the fictional representation of the African jungle. In particular, in Section 4.1, I will compare the representation of the jungle as “a place of darkness” across the TTs while, in Section 4.2, I will concentrate on the representation of the jungle as a personified entity in translation. My analysis will show that the alternation of the lexico-semantic patterns affects the cohesion of the translations, with consequences on the potential interpretation of text; in other words, I will illustrate the effects of alterations at the micro (textual) level on the macro (interpretational) level.

4.1 The representation of the jungle as a “place of darkness” in translation

This analysis follows the methodology used for the study of the ST: it starts from the examination of the concordance lines of the “Africa words” in order to check whether their translation reproduces the semantic preferences identified in the original. I will begin with *forest* and *jungle*, because these items have a direct Italian equivalent, *foresta* and *giungla* respectively, used by all the TTs. Table 5.5 and Table 5.6 collect the words forming part of the three semantic fields in the translations, in a span of 5:5.

Table 5.5 Collocates of *foresta* belonging to the three semantic fields

Translation S, <i>foresta</i>
ciglio (4)*, profondità (2), abbracciò, cinto, circondato, cupo, dentro, deserta, fosca, impenetrabile, lati, limite, misteriosamente, ombre, oscurità, pensosa, silenzio, spettrale, tenebroso, tetra.
Translation B, <i>foresta</i>
marginè (4), buia, circondata, circondato, cupo, deserta, entrata, impenetrabile, incupiva, mistero, ombre, profonda, profondità, selvaggi, selvaggia, silenzio, solitario, spettrale.
Translation G, <i>foresta</i>
ciglio (3), lato (2), profondità (2), abbracciava, circondata, circondava, cupo, dentro, espulso, immenso, impassibile, intrico, limitare, marginè, muraglia, neri, ombre, oscura, oscurità, selvaggia, silenzio, spettrale, tetro, usciva.
Translation M, <i>foresta</i>
cupo (2), marginè (2), orlo (2), profondo (2), silenzio (2), abbracciava, circondata, circondato, confuse, espulso, groviglio, impassibile, impenetrabile, lato, limite, limiti, nera, neri, ombre, parete, profondità, scuro, selvatica, silenzio, spaventoso, spettrale.

* Frequency, when ≥ 2 , in brackets

Table 5.6 Collocates of *giungla* belonging to the three semantic fields

Translation S, <i>giungla</i>
ciglio, colossale, cupo, immensa, misteriosamente, nere.
Translation B, <i>giungla</i>
intrico (2)*, marginè (2), fosco, immensa, immenso, invisibile, maledetta, muraglia, nere, profondità, selvaggi, usciva.
Translation G, <i>giungla</i>
colossale, immenso, impenetrabile, intrico, neri, selvaggi.
Translation M, <i>giungla</i>
colossale, immenso, impenetrabile, intrico, neri, orlo.

* Frequency, when ≥ 2 , in brackets

Tables 5.5 and 5.6 show that the semantic preferences are maintained in all four of the TTs. It is in fact possible to identify the words that create the dominant semantic fields in the original. *Foresta* and *giungla* occur with words conveying the idea of darkness (*cupo*, “gloomy”; *fosco*, “dull, dim”; *buio*, “dark”; *ombre*, “shadows”; *nere*, “black”; *mistero*, “mystery”; *spettrale*, “spectral”; etc.), with words conveying the idea of a closed and inaccessible place (*impenetrabile*, “impenetrable”; *marginè*, “margin”; *circondata*, “surrounded”; *limite*, “limit”; etc.), and with words conveying the idea of a place which is difficult to circumscribe due to its immensity and blurred boundaries (*profondità*, “depth”; *immense*, “immense”; *colossale*, “colossal”; *intrico*, “tangle”; etc.).

Moving on to *wilderness* and *darkness*, the analysis becomes less straightforward. These two words do not have a direct, one-to-one Italian equivalent, as in the case of *jungle* and *forest*. *Darkness* can have many different translations in Italian, depending on the context, whereas *wilderness* does not have a direct equivalent at all. The translators have to substitute or paraphrase them accordingly. Therefore, to compare these “Africa words” in translation, I have first to identify manually how they have been translated in each TT, by comparing the original text to the Italian versions. The outcome of this comparison is summarised in Table 5.7 below.

Table 5.7 Translations of *wilderness* and *darkness* for each TT

	<i>wilderness</i>	<i>darkness</i>
Translation S	terra selvaggia, “wild land” (6)* immensità selvaggia, “wild immensity” (3) mondo selvaggio, “wild world” (2) solitudine selvaggia, “wild solitude” (2) terra vergine, “virgin land” (2) deserto vegetale, “vegetation desert” foresta vergine, “virgin forest” profondità selvaggia, “wild depth” selvaggi recessi, “wild recesses” selvatica vita, “wild life” solitudine, “solitude” terra, “land” Total: 12 translations	tenebra, “darkness”, sing. (17) tenebre, “darkness”, plur. (5) oscurità, “darkness”/ “obscurity” tenebrosa aridità, “dark aridity” tenebroso, “dark”; adj. Total: 5 translations
Translation B	terra selvaggia, “wild land” (8) distesa desolata, “desolate expanse” (4) desolazione, “desolation” distesa selvaggia, “wild expanse” foresta, “forest” giungla, “jungle” natura selvaggia, “wild nature” plaga selvaggia, “wild region” selva, “woods” solitudine selvaggia, “wild solitude” solitudini selvage, “wild solitudes” terra desolata, “desolate land” Total: 12 translations	tenebra, “darkness”, sing. (20) tenebre, “darkness”, plur. (4) buio, “dark” Total: 3 translations

* Frequency, when ≥ 2 , in brackets

Table 5.7 (continued)

Translation G	landa selvaggia, “wild land” (13) selva selvaggia, “wild woods” (9)	tenebra, “darkness”, sing. (16) tenebre, “darkness”, plur. (5) oscurità, “darkness”/ “obscurity” (2) profondità tenebrosa, “dark depth” sterilità tenebrosa, “dark sterility”
	Total: 2 translations	Total: 5 translations
Translation M	terra selvaggia, “wild land” (17) immensità selvaggia, “wild immensity” natura selvaggia, “wild nature” regione selvaggia, “wild region” vita misteriosa e selvatica, “wild and mysterious life” zona selvaggia, “wild zone”	tenebre, “darkness”, plur. (14) tenebra, “darkness”, sing. (11)
	Total: 6 translations	Total: 2 translations

* Frequency, when ≥ 2 , in brackets

Translation S adopts 12 different translations for *wilderness* and 5 for *darkness*; Translation B, 12 for *wilderness* and 3 for *darkness*; Translation G translates *wilderness* and *darkness* in 2 and 5 different ways respectively, while Translation M uses 6 for *wilderness* and 2 for *darkness*. Once these translations have been identified, I can analyse their concordance lines so as to check whether the dominant semantic fields are also maintained in Italian or not. Consider Table 5.8 and Table 5.9 below, which list the “impenetrableness”, “darkness”, and “immensity” words that occur with the Italian translations of *wilderness* and *darkness* in a span of 5:5.

Table 5.8 Collocates of the translations of *wilderness* belonging to the three semantic fields

Translation S, Italian translations of <i>wilderness</i>
abbandonata, attorno, misteriosamente, mistero, muta, muto, racchiudeva, richiuse, perduto, silenziosa, sola, vuota.
Translation B, Italian translations of <i>wilderness</i>
abbandonata, addentrava, chiuse, circondava, colossale, desolata, dolente, enorme, greve, immensa, misteriosa, mistero, muto, profondità, riaccolse, silente, sola, sperduto, vuoto.
Translation G, Italian translations of <i>wilderness</i>
silenziosa (2)*, circondava, colossale, dimenticata, greve, immensa, insondabile, isolamento, misteriosa, mistero, muto, profondità, richiuse, sperduto, vuota.
Translation M, Italian translations of <i>wilderness</i>
chiusa, circondava, desolata, dimenticata, mistero, muto, perduto, profondità, richiuse, silenziosa, vuota.

* Frequency, when ≥ 2 , in brackets

Table 5.9 Collocates of the translations of *darkness* belonging to the three semantic fields

Translation S, Italian translations of <i>darkness</i>
impenetrabile (2)*, impenetrabile, inabissava, occulta, pericolo, porta, profonda, profondamente, profondo, soglia, spiraglio.
Translation B, Italian translations of <i>darkness</i>
impenetrabile (3), aperta, immensa, nera, pericolo, porta, profonda, profondità, soglia.
Translation G, Italian translations of <i>darkness</i>
impenetrabile (3), circondava, immensa, invisibile, nascosto, profonda, socchiuso, soglia.
Translation M, Italian translations of <i>darkness</i>
impenetrabile (2), porta (2), aperta, immensa, impenetrabili, infittivano, invisibile, nascosto, nera, profonda, profondità.

* Frequency, when ≥ 2 , in brackets

All four TTs maintain the semantic fields identified in the original, both in relation to *wilderness* and *darkness*. The Italian versions of *wilderness* and *darkness* equally co-occur with terms conveying the idea of “impenetrableness” (*racchiudeva*, “enclosed”; *richiuse*, “closed”; *circondava*, “surrounded”; *impenetrabile*, “impenetrable”; *porta*, “door”; *soglia*, “threshold”; etc.), “darkness” (*misteriosamente*, “mysteriously”; *silenziosa*, “silent”; *dolente*, “sorrowful”; *mistero*, “mystery”; *occulta*, “occult”; *pericolo*, “danger”; *nera*, “black”; etc.), and “immensity” (*perduto*, “lost”; *colossale*, “colossal”; *enorme*, “enormous”; *profondità*, “depths”; *immensa*, “immense”; *sperduto*, “lost”; *profonda*, “profound”; etc.).

Considering also the collocates of *foresta* and *giungla*, it can be noted that the repetition of the exact same terms and forms of the same lemmas across the “Africa words” is a feature of the Italian semantic fields as well. In Translation S, *mistero/misteriosamente* (“mystery”/“mysteriously”), *silenzio/silenziosa* (“silence”/“silent”), *impenetrabile* (“impenetrable”), *ciglio* (“edge”), *cupo* (“gloomy”), and *profonda/profondamente/profondo* (“deep”/“deeply”/“deep”) are shared by two or more “Africa words”. In Translation B, *silente/silenzio* (“silent”/“silence”), *immensa/immenso* (“immense”), *selvaggi* (“wild”), *circondava/circondata/circondato* (“surrounding”/“surrounded”/“surrounded”), *impenetrabile* (“impenetrable”), *nera/nere* (“black”), *profondità/profonda* (“depth”/“deep”), and *mistero/misteriosa/misteriosamente* (“mystery”/“mysterious”/“mysteriously”) are shared by two or more “Africa words”. In Translation G, *silenziosa/silenzio* (“silent”/“silence”), *circondava/circondata* (“surrounding”/“surrounded”), *immensa/immenso* (“immense”), *profondità/profonda*

(“depths”/“deep”), *impenetrabile* (“impenetrable”), *intrico* (“tangle”), *neri* (“black”), *selvaggia/selvaggi* (“wild”), and *colossale* (“colossal”) are shared by two or more “Africa words”. Finally, in Translation M, *nera/neri* (“black”), *circondava/circondata/circondato* (“surrounding”/“surrounded”/“surrounded”), *silenziosa/silenzio* (“silent”/“silence”), *impenetrabile/impenetrabili* (“impenetrable”), *immensa/immenso* (“immense”), and *profondità/profonda/profondo* (“depths”/“deep”) are shared by two or more “Africa words”.

Summing up, the Italian translations of *forest*, *wilderness*, *jungle*, and *darkness* establish equivalent lexico-semantic patterns as in the original, enacting a comparable local textual function. Generally speaking, the fictional representation of the African jungle as a “place of darkness” is identifiable in all four TTs. However, what is irremediably altered in the translations is the lexical cohesion these patterns establish in the original.

4.2 Lexical cohesion in the translations

In Section 2, I have argued that the persistent use of the “impenetrableness”, “darkness”, and “immensity” semantic fields with all four of the “Africa words” establishes a lexical network that makes the fictional representation of the African jungle cohesive. In the previous section, I have shown that the TTs reproduce the semantic preferences of the original, keeping the network between the Italian “Africa words” unaltered. However, in all four of the TTs, *wilderness* and *darkness* have been translated in multiple ways: in this section, I will demonstrate that this choice affects significantly the lexical cohesion of the TTs.

Even though the semantic fields are there, the lexical network that connects the “Africa words” to these fields is less tight, because in the TTs the “Africa words” are more numerous than in the original. The use of multiple translations for the same ST item loosens the cohesive chains and dilutes the lexico-semantic pattern which, in the ST, is based on fewer items. Consider the case of *wilderness*, for instance. *Wilderness* is the “Africa word” that has been translated in the most different ways, so it lends itself to explaining the loss of cohesion in the TTs. In the ST, every time *wilderness* occurs, the reader finds in its surrounding context similar words and

semantic fields which, reiterated throughout the whole text, enact a cohesive and consistent image of the wilderness. Example (5) below exemplifies this phenomenon: six sentences in which *wilderness* occurs are gathered and put together one after the other, in the order they appear in the text. The node word is in bold while the “impenetrableness”, “darkness”, and “immensity” words are in italics. Every time *wilderness* occurs, the semantic preference is reiterated and a new tile is added to the overall mosaic to create a cohesive fictional representation.

- (5) Land in a swamp, march through the woods, and in some inland post feel the savagery, the utter *savagery*, had *closed* round him – all that *mysterious* life of the **wilderness** that stirs in the forest, in the jungles, in the hearts of wild men. [...] And *outside*, the *silent wilderness* surrounding this cleared speck on the earth struck me as something great and invincible, like *evil* or truth, waiting patiently for the passing away of this fantastic invasion. [...] In a few days the Eldorado Expedition went into the patient **wilderness**, that *closed* upon it as the sea closes over a diver. [...] And, after all, they did not eat each other before my face: they had brought along a provision of hippo-meat which went rotten, and made the *mystery* of the **wilderness** stink in my nostrils. [...] But the **wilderness** had found him out early, and had taken on him a *terrible vengeance* for the fantastic invasion. [...] And in the hush that had fallen suddenly upon the whole *sorrowful* land, the *immense wilderness*, the *colossal* body of the fecund and *mysterious* life seemed to look at her, pensive, as though it had been looking at the image of its own *tenebrous* and passionate soul.

In the first sentence, *closed* and *mysterious* occur in relation to *wilderness*. Later in the novel, the reader finds *wilderness* co-occurring with *closed* again, in the third sentence. *Mystery*, in the fourth sentence, recalls the previous use of *mysterious*, while a verbatim repetition of *mysterious* reappears again in the context of *wilderness* in the

sixth sentence. This lexical network, spread consistently throughout the whole text, establishes a cohesive representation of the wilderness.

In translation, the link that connects one occurrence of *wilderness* to the other is interrupted because the term has been rendered with different items. Even though the semantic fields are there, they are not tied throughout the whole text since the reiterated node word which acts as the spine of the lexical network is missing. Example (5S) shows the same sentences in Example (5) from Translation S, selected because its wide range of different translations of *wilderness* is exemplificatory.

(5S) Sbarcare in un pantano, marciare attraverso i boschi, e in qualche posto avanzato dell'interno sentire che la natura *selvaggia*, tutto quel che si può dare di più *selvaggio* s'è *richiuso* attorno a lui; tutta quella **selvatica vita** che si agita *misteriosamente* nella giungla, nella foresta, nel cuore dei barbari. [...] E, al di *fuori*, quella *silenziosa* **immensità selvaggia** che *racchiudeva* quel minuscolo spiazzo sulla terra m'appariva non so che di grande e d'invincibile, come il *male* o il vero, pazientemente in attesa della fine di quella grottesca invasione. [...] Di lì a pochi giorni la «Spedizione Eldorado» s'inabissò in quel paziente **deserto vegetale**, che si *richiuse* su di lei come il mare sopra un palombaro. [...] E poi, dopo tutto, non si mangiarono mai l'un l'altro in mia presenza. S'eran portata seco una provvista di carne d'ippopotamo che finì per imputridire, e mi rese sensibile il *mistero* di quella **terra vergine** in un fetore che mi empiva le narici. [...] Ma quella **terra selvaggia** lo aveva capito assai presto, e aveva preso su di lui una *tremenda vendetta* per tutta quella grottesca invasione. [...] E nel silenzio che era *subitaneamente* caduto su tutta quella *accorata* contrada, l'**immensità selvaggia**, il corpo *colossale* di quella vita feconda e *misteriosa* pareva la guardasse, intenta, quasi contemplando l'immagine della propria anima, *tenebrosa* e appassionata.

“Disembarking in a morass, marching through the woods, and in some place inland feeling that the *wild* nature, all that can be most *wild*, has *closed* round him; all that **wild life** that stirs *mysteriously* in the jungle, in the forest, in the heart of the barbarians. [...] And, *outside*, that *silent* **wild immensity** that *closed* that minuscule open space on the earth appeared to me as something great and invincible, like *evil* or truth, waiting patiently the end of that grotesque invasion. [...] In a few days the “Eldorado Expedition” sank into that patient **vegetation desert**, that *closed* upon her as the sea upon a diver. [...] And then, after all, they never ate each other in my presence. They had brought with them a provision of hippo-meat which ended up rotting, and made me sensitive to the *mystery* of that **virgin land** with a fetor that filled my nostrils. [...] But that **wild land** had understood it very soon, and had taken on him a *terrible vengeance* for all that grotesque invasion. [...] And in the *silence* that had fallen suddenly upon that earnest region, the **wild immensity**, the *colossal* body of that fecund and *mysterious* life seemed to look at her, intent, as though staring at the image of her own souls, tenebrous and passionate.”

Although the semantic preference is identifiable, the link between one instance of it and the following one is looser, as the semantic fields refer, in this specific example, to five different node words.

In the first sentence of Example (5S), *misteriosamente* (“mysteriously”) and *richiuso* (“closed”) occur in the context of *selvatica vita* (“wild life”), as in the original. In the third sentence, *richiuse* (“closed”) occurs again, recalling the previous use of *richiuso* (“closed”). However, this time *richiuse* (“closed”) refers to *deserto vegetale* (“vegetation desert”), and not anymore to *selvatica vita* (“wild life”). Similarly, the occurrences of *mistero* (“mystery”) in the fourth sentence and *misteriosa* (“mysterious”) in the sixth sentence build on each other, as in the ST, but in this case too the link that relates them is missing. In fact, *mistero* (“mystery”) and *misteriosa* (“mysterious”) refer to two different nodes, *terra vergine* (“virgin land”)

and *immensità selvaggia* (“wild immensity”). The lexical network is therefore interrupted.

Baker (2011: 215) acknowledges that, when there is a lack of ready equivalents, subtle or even major changes are sometimes inevitable in translation. If these changes affect the cohesion of the text, then Baker (2011) suggests substituting the lexical chain of the ST with a different but equivalent chain in the TT, as long as lexical cohesion is established. On the contrary, what the translator must always avoid is “the extreme case of producing what appears to be a random collection of items which do not add up to recognizable lexical chains that make sense in a given context” (Baker 2011: 216). This is what seems to happen in Translation S, where the lexical chain is interrupted and substituted by “a random collection of items” (the different translations of *wilderness*) that “do not add up” to the same “recognizable lexical chains” (the repetition of the core word *wilderness* in the ST). The lexical chain that links the instantiations of the semantic preferences is established only when *wilderness* is translated with the same expression, for instance, *immensità selvaggia* (“wild immensity”) in Example (5S). As a result, the cohesion of the wilderness is less tight than in the ST.

Another consequence of translating *wilderness* with multiple terms, which equally affects the lexical cohesion of the TTs, is the loss of emphasis: the TT items (the Italian translations of *wilderness*) do not carry the same emphasis that the ST item (*wilderness*) carries. This loss of emphasis is due to the strategy that is usually adopted to translate *wilderness* in Italian, namely, paraphrasing. Paraphrasing has the advantage of achieving a high level of precision in conveying the semantic meaning of the original item. However, it has the disadvantage of filling one-item slot with a longer multi-item expression (Baker 2011: 41). The most common way the noun *wilderness* is translated in the TTs is by paraphrasing it with the construction “noun + adjective”, as in *terra selvaggia* (“wild land”) or *natura selvaggia* (“wild nature”). *Selvaggio/selvaggia* (“wild”) is the most common adjective used in this construction, which carries most of the semantic meaning of *wilderness*. As a result, the emphasis that *wilderness* carries as a one-item noun is spread over two elements, a newly introduced noun and an adjective post- or pre-modifying it.

This strategy alters the role *wilderness* plays in establishing cohesion. Although the semantic meaning of *wilderness* is to some extent reproduced by

selvaggio/selvaggia (“wild”), “expressive, evoked or any kind of associative meaning[s]” (Baker 2011: 41) cannot be equally maintained through paraphrasing. These meanings are “associated only with stable lexical items which have a history of recurrence in specific contexts” (Baker 2011: 41). In the ST, *wilderness* does have such a “history of recurrence in specific contexts”, i.e. semantic preferences, in corpus linguistics terms. *Wilderness* co-occurs recurrently within the same specific semantic fields. This relation between the core *wilderness* and its contexts of occurrence provides *wilderness* with extra meanings, in addition to the semantic one: the expressive, evoked, and associative meanings that Baker (2011: 41) refers to. Paraphrasing disrupts the relation between the core and its contexts of occurrence, because it substitutes a stable lexical item in the ST (*wilderness*) with multiple items that do not recur consistently enough in TTs. Some instances of this phenomenon can be found in Example (6) and Example (7) below:

(6) But the *wilderness* had found him out early, and had taken on him a terrible vengeance for the fantastic invasion.

(6S) Ma quella *terra selvaggia* lo aveva capito assai presto, e aveva preso su di lui una tremenda vendetta per tutta quella grottesca invasione.

“But that *wild land* had understood him very early, and had taken on him a terrific vengeance for all that grotesque invasion.”

(7) And outside, the silent *wilderness* surrounding this cleared speck on the earth struck me as something great and invincible, like evil or truth, waiting patiently for the passing away of this fantastic invasion.

(7B) E fuori, la silente *distesa desolata* che circondava questo bruscolo disboscato sulla terra mi colpiva come qualcosa di grande e invincibile, come il male o la verità, pazientemente in attesa che scomparisse questa fantastica invasione.

“And outside, the silent *desolate expanse* that surrounded this cleared speck on the earth struck me as something great and invincible, like evil or truth, waiting patiently for this fantastic invasion to disappear.”

In Examples (6S) and (7B) the emphasis of *wilderness* is shifted to the noun *terra* (“land”) and *distesa* (“expanse”) respectively, introduced anew by the translators. In the case of Example (6S), although the semantic meaning of *wilderness* is maintained by the post-modifying adjective *selvaggia* (“wild”), the emphasis that *wilderness* carries in the ST is shifted on a TT item (*terra*, “land”) which does not have the same “history of recurrence in specific contexts” (Baker 2011: 41) as *wilderness*. In the case of Example (7B), the semantic meaning of *wilderness* is lost altogether, since the adjective post-modifying *distesa* (“expanse”) has a completely different semantic meaning from *wilderness*, namely *desolata* (“desolate”). As a result, the emphasis the noun *wilderness* carries as a subject is moved to a TT item which has a different semantic meaning and interpretational implications. Overall, the nouns introduced by the translators can be regarded as hyponyms of the original item: *terra* (“land”) and *distesa* (“expanse”) can be seen as more specific terms than *wilderness*. This alteration narrows down the metaphorical and symbolic implications of the vaguer term *wilderness*, through an Italian translation that is more concrete and precise in conveying the idea of an actual place, instead of a figurative space, a metaphor, as it happens in the original. This is a significant alteration, considering that HoD plays repeatedly with the idea of vagueness.

These two types of alterations (disruption of the lexical chain and loss of emphasis) affect irremediably the local textual functions that the “Africa words” enact in the original. In turn, the lexical cohesion they establish is affected too. These alterations do not involve *forest* and *jungle*, as they have a direct Italian equivalent with which they have been consistently translated. Nevertheless, I have shown that the fictional representation of the African jungle hinges on the interrelation of all of the

four “Africa words” together. Therefore, even the alterations involving *wilderness* and *darkness* only have a major impact on the overall depiction of the jungle. As a consequence, the interpretational implications that the depiction of the African jungle raises in the ST may be less evident in the TTs, missing consistency and cohesion. In other words, the target reader can perceive the theme “Africa and its representation” differently from the source reader, because the construction of the African jungle in the translation is different from that of the original.

It must be clarified though that the alterations I have discussed in this section affect the different translations to different extents. Translation S is the TT with the highest number of different variations for *wilderness* and *darkness*. Supposedly, this is the Italian version in which the depiction of the jungle is most heavily altered. At the other end of the scale, there is Translation G which, although it presents 5 different translations of *darkness*, translates *wilderness* with only 2 variants. In this case, keeping the same terms throughout the text helps creating a lexical network which is more cohesive than that created by the other translations. In Translation G, the semantic preferences are established around only two variants instead of 12, as in the case of Translation S.

4.3 Personification in translation

In Section 3, I have argued that specific lexical choices personify the African jungle. *Forest* and *wilderness* operate as active agents that perform physical actions, possess cognitive functions, and display a malevolent disposition towards external intruders. This malicious characterisation is also manifested in terms of a seductive yet dangerous woman, charming and wicked at the same time. Building on the discussion of the translation of *forest* and *wilderness* in the previous section, in this section I will extend the comparison of the TTs by investigating what happens to the personification tendency, whether and how it has been maintained in translation.

Figure 5.6 below shows the Italian concordance lines of *foresta*, the same concordance lines of *forest* showed in Figure 5.5.

Translation S

1	quelle loro doghe. Credo davvero che se le portassero in letto. Di là dallo steccato la foresta s'ergeva spettrale nel chiaro di luna, e attraverso al rimescolio confuso, alle fievoli
2	senza che fosse percettibile alcun movimento di ritirata, non altrimenti che se la foresta, la quale aveva buttato fuori quelle creature tanto improvvisamente, le avesse
3	nuovo tratto di fiume si apriva innanzi a noi, e si chiudeva alle nostre spalle, come se la foresta avesse tranquillamente attraversato le acque dall'una all'altra sponda per sbarrarci
4	scudi, con sguardi feroci e movimenti selvaggi, si rovesciarono nella radura fuori da quella foresta fosca e pensosa. I cespugli si agitarono, l'erba ondeggiò un tratto, e poi ogni cosa
5	mattì su quel ponte di ferro. Un frastuono tremendo usciva fuori da quella carcassa, e la foresta vergine dall'altra parte della cala lo rifrangeva in un tonante rombo sulla stazione
6	qualcosa di così stupefacente che io balzai in piedi e mi volsi a guardare verso il limite della foresta, come se m'aspettassi un qualche segno di risposta a quella diabolica ostentazione
7	riuniti i paletti tra loro, era scomparsa. Tutto ciò era, naturalmente, circondato dalla foresta. Di sotto alla casa la ripa era parimenti sgombra, e proprio in riva all'acqua vidi un
8	si proiettava diritta sopra al fiume. Il pomeriggio era già alquanto inoltrato, il volto della foresta era cupo, e un'ampia striscia d'ombra era già scesa sulle acque. Si avanzava

Translation B

1	I tratti di fiume si aprivano davanti a noi per chiudersi alle nostre spalle come se la foresta fosse entrata nell'acqua a passi indolenti per sbarrarci la via del ritorno.
2	loro bastoni in mano. Penso davvero che se li portassero a letto. Oltre il recinto la foresta si dispiegava spettrale al chiaro di luna, e attraverso l'attutto tramestio,
3	capiole sul ponte di ferro. Un dangore spaventoso si levò da quello scafo, e la foresta vergine, sulla riva opposta della caletta, lo rimandò indietro con un rombo di
4	dei selvaggi si diradava senza alcun percettibile movimento di ritirata, come se la foresta, che così all'improvviso aveva eruttato quegli esseri, li avesse risucchiati
5	. Era così sorprendente che balzai in piedi e mi voltai a guardare il margine della foresta, quasi che mi fossi atteso una risposta qualsiasi a quel nero sfoggio di
6	c'era stato tra i pali, erano scomparse. Naturalmente, il tutto era circondato dalla foresta. La riva era disboscata, e ai bordi dell'acqua vidi un bianco sotto un
7	si protendeva rigido sulla corrente. Era ormai pomeriggio inoltrato, il volto della foresta s'incupiva e una larga striscia d'ombra era già caduta sull'acqua. In
8	sguardi feroci e movimenti selvaggi, si riversarono dalla facciata buia e triste della foresta sulla radura. I cespugli si agitavano, l'erba ondeggiò per un certo tempo e

Translation M

1	a saltellare sul ponte di ferro. Dallo scafo si alzò un frastuono spaventoso e la foresta vergine sull'altra riva dell'insenatura lo rimandò come un rombo tonante
2	Sono convinto che li portassero a letto con loro quei bastoni. Oltre lo steccato la foresta si ergeva spettrale al chiaro di luna e attraverso la confusa agitazione,
3	più grosso si proiettava rigido sulla corrente. Era ormai pomeriggio inoltrato, la foresta aveva un volto cupo e un'ampia striscia d'ombra era già calata
4	folia di selvaggi svaniva senza alcun percettibile movimento di ritirata, come se la foresta che aveva espulso quegli esseri all'improvviso, li avesse risucchiati come
5	una freccia aguzza diretta al cuore stesso della terra; e come per incanto, la foresta dal volto scuro e pensoso riversò nella radura fiumi di esseri umani – esse
6	di fiume si aprivano davanti a noi e si richiudevano alle nostre spalle come se la foresta attraversasse con tutto comodo l'acqua per sbarrarci la via del ritorno.
7	avesse collegati, erano scomparse. Naturalmente tutto questo era circondato dalla foresta. La sponda del fiume era sgombra, e presso l'acqua vidi un bianco sotto
8	impressionante che balzai in piedi per guardare, alle mie spalle, il margine della foresta, come se mi aspettassi una risposta qualunque a quell'oscura

Translation G

1	si aprivano davanti a noi e si richiudevano al nostro passaggio, come se la foresta, pigra e tranquilla, avesse scavalcato l'acqua per sbarrarci la via del ritorno
2	. Credo proprio che se li portassero a letto, quell'arnese. Oltre la staccionata la foresta si ergeva spettrale al chiaro di luna, e attraverso il fermento indistinto,
3	di ferro. Dallo scafo smantellato uscì uno spaventoso rumore di ferraglia che la foresta vergine, dall'altro lato dell'insenatura, rimandò come un rombo di tuono
4	si stava disperdendo senza alcun percettibile movimento di ritirata, come se la foresta che aveva espulso quelle creature così all'improvviso, ora le risucchiasse,
5	, o quello che poteva esserci in mezzo a loro, erano sparite. Naturalmente la foresta circondava tutto, ma la riva era sgombra e sul bordo dell'acqua vidi un
6	terra. Era così stupefacente che balzai in piedi e mi voltai a guardare il ciglio della foresta, quasi mi aspettassi una qualche risposta a quel diabolico sfoggio di
7	ramo si protendeva rigido di traverso. Nel pomeriggio ormai inoltrato, il volto della foresta appariva cupo, e sull'acqua era già scesa una larga striscia d'ombra. Era
8	, archi e scudi, con sguardi feroci e movimenti selvaggi, si riversò nella radura dalla foresta dal volto scuro e pensoso. La boscaglia fremette, l'erba ondeggiò un

Figure 5.6 Concordance lines of *foresta*: personification

Overall, all four of the TTs maintain the personification associated with *foresta* (“forest”), mainly as a result of a literal translation of Conrad’s lexical choices. For example, the forest is *vergine* (“virgin”), it has a *volto cupo* (“dark face”, apart from Translation B where *il volto si incupiva*, “the face darkened”) and *pensono* (“pensive”), and it is expected to give a *risposta* (“answer”) to Marlow in all of the four TTs. However, in other cases the translators level out the personification, replacing some of Conrad’s lexical choices with items that fit more standardly the description of a geographical landscape. This is the case of *the forest stood up* (line 1 in Figure 5.5). In Translations S, M, and G, it becomes *la foresta si ergeva* (“the

forest rose”) and in Translation B it becomes *la foresta si dispiegava* (“the forest spread”). Similarly, in the case of *the forest surrounded all that* (line 6 in Figure 5.5), *foresta* (“forest”) loses its role as the subject of the sentence in three out of four TTs, becoming instead the agent of a passive construction. In Translations S, B, and M, *the forest surrounded all that* becomes *tutto ciò era, naturalmente, circondato dalla foresta* (“all of that was, naturally, surrounded by the forest”), *il tutto era circondato dalla foresta* (“all of that was surrounded by the forest”), and *tutto questo era circondato dalla foresta* (“all of this was surrounded by the forest”), respectively. The forest is deprived of its active agency, while the impression that it is performing a physical action is toned down, in accordance with a more standardised representation of the African jungle as an inanimate entity. On the contrary, Translation G maintains the active construction with *foresta* as subject: *la foresta circondava tutto* (“the forest surrounded all that”).

With regard to *wilderness*, the lack of a one-to-one equivalent in Italian adds a further level of complexity to the analysis. Consider Figure 5.7 below, which shows the different translations of *wilderness* in the same concordance lines showed in Figure 5.5 above. The Italian translations of *wilderness* too maintain overall the personifying tendency. For instance, in all four TTs, the wilderness paws (*zampate*, “blow with a paw”), laughs (*prodigiosa risata*, “prodigious laugh”, line 9 in Translation S; *enorme risata*, “enormous laugh”, line 9 in Translation B; *formidabile risata*, “formidable laugh”, line 11 in Translation M; *fragorosa risata*, “uproarious laugh”, line 11 in Translation G), and seems to be looking at Kurtz’s mistress (*pareva la guardasse*, “seemed that it looked at her”, line 5 in Translation S; *pareva guardarla*, “appeared to look at her”, line 6 in Translation B; *sembrava [...] la guardasse*, “seemed that it looked at her”, line 1 in Translation M; *sembrava pensosamente guardarla*, “seemed it was looking at her pensive”, line 2 in Translation G). The attribution of a breast, bosom, and fecundity to the wilderness is also maintained in all the TTs: *breast* in line 1 of Figure 5.5 becomes *seno* (“bosom”) in Translations S and B, *petto* (“breast”) in Translation M, but *cuore* (“heart”) in Translation G; *bosom* in line 2 of Figure 5.5 becomes *seno* (“bosom”) in Translations S, B, and M, and *grembo* (“lap”/“womb”) in Translation G; finally, *fecund* in line 8 of Figure 5.5 is translated as *feconda* (“fecund”) in all four TTs.

Translation S

1	pur curvarne un filo. Di lì a pochi giorni la «Spedizione Eldorado» s'inabissò in quel paziente deserto vegetale, che si richiuse su di lei come il mare sopra un palombaro. Dopo molto tempo
2	rispetto a qualsiasi fine pratico. Cerca di rompere l'incanto — il greve e muto incanto della foresta vergine — che pareva se lo attrasse al seno, risvegliando in lui certi istinti brutali e
3	. Per Giove! Non ho mai visto nulla di tanto irreali in vita mia. E, al di fuori, quella silenziosa immensità selvaggia che racchiudeva quel minuscolo spiazzo sulla terra m'appariva non so che
4	sofferente, stava cercando di riaversi: finalmente s'alzò e se n'andò via, e quella muta immensità selvaggia lo accolse di nuovo nel suo seno. Uscendo dall'oscurità, m'andavo
5	incedere. E nel silenzio che era subitamente caduto su tutta quella accorata contrada, l'immensità selvaggia, il corpo colossale di quella vita feconda e misteriosa pareva la guardasse
6	seguì. «Oh, ma io avrò ancora modo di strapparti il cuore», egli gridò verso quell'invisibile mondo selvaggio. Un bel giorno la macchina s'arrestò — come avevo pur previsto — e ci
7	il batter d'un cuore — il cuore di una tenebra vittoriosa. Era un momento di trionfo per quella terra selvaggia, una irruzione veemente e vendicatrice che, a quanto mi sembrava, mi sarebbe
8	abbia finito con l'averne piena coscienza da ultimo: proprio solo da ultimo, però. Ma quella terra selvaggia lo aveva capito assai presto, e aveva preso su di lui una tremenda vendetta
9	gli apparteneva. E io trattenevo senso volerlo il respiro nell'attesa di udir scoppiare quella terra selvaggia in una prodigiosa risata che avrebbe fatto tremare le stelle nelle loro sedi. Ogni
10	di crescere: ma questo, come dire, questo esemplare era di una calvizie impressionante. Quella terra selvaggia gli aveva dato un buffetto sul capo: ed eccola di colpo, la sua testa simile a
11	spesso «un po' di febbre» o un po' di questo o di quell'altro — scherzose zampate di quella terra selvaggia, frigoli preliminari di quel più serio assalto che venne poi a suo tempo. Dunque,

Translation B

1	cadavere. Per Giove! Non ho mai visto niente di più irreali in vita mia. E fuori, la silente distesa desolata che circondava questo bruscolo disboscato sulla terra mi colpiva come
2	piegarne un filo. — Di lì a pochi giorni la Spedizione Eldorado si addentrava nella paziente distesa desolata, che si chiuse su di essa come il mare su di un tuffatore. Molto tempo
3	con un aspetto molto sofferente, a cercare di riprendersi: poi si alzò e se ne andò e la distesa desolata senza un suono lo riaccolse nel suo seno. Mentre dal buio mi avvicinavo al
4	spettacolo. » Lo feci. Ci fu silenzio. «Oh, ma saprò ancora strapparti il cuore!», gridò alla giungla invisibile. — Come avevo previsto, andammo in avaria e si dovette stare all'ancora
5	a crescere, ma questo... ah!... questo esemplare era calvo da fare impressione. Quella terra selvaggia gli aveva dato un colpo sulla testa e, guardate, era diventata come una
6	lento. E nel silenzio discesi all'improvviso su tutta quella plaga dolente, l'immensa terra selvaggia, quel regno colossale di vita feconda e misteriosa pareva guardarla, pensoso
7	. Avevo spesso «un po' di febbre», o un po' di qualcos'altro; giucose zampate della terra selvaggia, il trascurabile annuncio del più serio attacco che arrivò a suo tempo. Sì, li
8	utilità a ogni fine pratico. Tentavo di rompere l'incantesimo — il greve, muto sortilegio della terra selvaggia, — che sembrava lo attraesse nel suo seno spietato risvegliando istinti
9	mio...»; tutto gli apparteneva. Mi faceva tenere col fiato in sospenso aspettando di udire la terra selvaggia scoppiare in una enorme risata che avrebbe scosso le stelle fisse nelle loro
10	il battito di un cuore, il cuore di una tenebra vincitrice. Fu un momento di trionfo per la terra selvaggia, un'irruzione dilagante e vendicativa che, mi parve, avrei dovuto respingere
11	di dirlo. Credo che la consapevolezza gli sia arrivata alla fine, proprio solo alla fine. Ma la terra selvaggia lo aveva scoperto presto, e si era presa su di lui una terribile vendetta per

Translation M

1	nel suo incedere deliberato. E nel silenzio che era caduto improvviso su quella terra addolorata, sull'immensità selvaggia, sembrava che quel corpo colossale dalla vita misteriosa e feconda la guardasse
2	d'ombra con l'aria molto sofferente, che cercava di riprendersi; quindi si alzò e se ne andò — e la natura selvaggia se lo riprese in seno senza il minimo rumore. Mentre dal buio mi avvicinavo al fuoco, mi
3	conto di quel suo difetto. Credo che alla fine debba averne avuto coscienza — soltanto alla fine. Quella terra selvaggia, però, l'aveva scoperto presto, e si era vendicata su di lui in modo terribile della
4	Avevo sovente «un po' di febbre» o un accenno di qualche altro malessere — zampate giucose di quella terra selvaggia, avvisaglie dell'assalto più serio che sarebbe arrivato a suo tempo. Sì; li osservavo
5	la minima utilità pratica. Cercavo di spezzare l'incantesimo — il pesante, muto incantesimo di quella terra selvaggia — che sembrava attrarlo al suo petto impietoso risvegliando dimenticati istinti brutali,
6	, senza piegarne neppure un filo. «Pochi giorni dopo la Spedizione Eldorado se ne andò nella paziente terra selvaggia e questa si richiuse su di essa come il mare si richiude su di un tuffatore. Molto tempo
7	.» Lo chiusi. Ci fu un attimo di silenzio. «Oh, ma ti strapperò ugualmente il cuore!» gridò all'invisibile terra selvaggia. «Avevamo un'avaria — come avevo previsto — dovemmo ancorarci alla punta di un'isola
8	i capelli seguitano a crescere, ma questo — ah — esemplare era calvo in modo impressionante. La terra selvaggia gli aveva accarezzato la testa ed eccolo lì, calvo come una palla — una palla d'avorio; io
9	zaffata da qualche cadavere. Per Giove! Non ho mai visto nulla di tanto irreali in vita mia. E fuori, la terra selvaggia e silenziosa che circondava quella minuscola radura mi pareva grande e invincibile quanto
10	e smorzato come il battito del cuore — il cuore di una tenebra vittoriosa. Fu un attimo di trionfo per la terra selvaggia, un'irruzione pervasiva e vendicatrice che, mi parve, avrei dovuto respingere da solo
11	la mia stazione, il mio fiume, il mio...» tutto gli apparteneva. Io trattenevo il fiato in attesa di udire la terra selvaggia scoppiare in una formidabile risata che avrebbe scosso le stelle fisse dalla loro sede. Gli

Translation G

1	alta senza piegarne un solo filo. «Di lì a pochi giorni la Spedizione Eldorado si inoltrò nella paziente landa selvaggia, che si richiuse su di lei come fa il mare sopra uno che si tuffa. Dopo molto tempo
2	nel suo incedere risoluto. E nell'improvviso silenzio caduto su quella terra afflitta, l'immensa landa selvaggia, quel corpo colossale di vita feconda e misteriosa sembrava pensosamente guardarla
3	. Per Giove! Non ho mai visto niente di tanto irreali nella mia vita. E intorno, la silenziosa landa selvaggia che circondava quel pezzetto disboscato di terra, mi colpiva come qualcosa di grande
4	sofferente, mentre stava cercando di riprendersi; finalmente si alzò e se ne andò, e la silenziosa landa selvaggia se lo riprese in grembo. «Mentre mi avvicinavo al bagliore provenendo dall'oscurità,
5	. Avevo spesso «un po' di febbre» o un leggero attacco di altre cose: le zampate scherzose della landa selvaggia, le iniziali schermaglie che precedono l'assalto più serio che venne poi a tempo
6	a crescere, ma questo... ehm... questo esemplare era di una calvizie impressionante. La selva selvaggia gli aveva dato un buffetto sulla testa, ed ecco, era diventata come una palla: una
7	quello che chiedeva. Ci fu silenzio. «Oh, ma te lo strapperò il cuore, vedrai!», gridò all'invisibile selva selvaggia. «Ci fu un'avaria — come mi ero aspettato — e dovemmo fermarci sulla punta di
8	servito praticamente a nulla. Cercavo di rompere l'incantesimo — il greve, muto incantesimo della selva selvaggia — che sembrava volerlo attrarre nel suo cuore impietoso risvegliando istinti brutali e
9	come il battito di un cuore — il cuore di una tenebra vittoriosa. Fu un momento di trionfo per la selva selvaggia, un'incurisione invadente e vendicativa che a me sembrava di dover respingere da
10	deficienza, io non lo so. Credo che se ne sia reso conto alla fine, quasi all'ultimo istante. Ma la selva selvaggia lo aveva scovato subito, e si era presa una terribile vendetta su di lui per quella
11	mia stazione, il mio fiume, il mio...» Era tutto suo. E io trattenevo il fiato aspettandomi di udire la selva selvaggia scoppiare in una fragorosa risata che avrebbe scosso le stelle fisse sul loro asse.

Figure 5.7 Concordance lines of the translations of *wilderness*: personification

Summing up, it can be said that the African jungle is a personified entity in all four TTs, as seen in the concordance lines of both *foresta* and of the Italian translations of *wilderness*. It performs behavioural and mental processes, laughing, looking, and finding out (*lo aveva capito*, “had understood it”, line 8 in Translation S; *lo aveva scoperto*, “had found it out”, line 11 in Translation B; *l'aveva scoperto*, “had found it

out”, line 3 in Translation M; *lo aveva scovato*, “had tracked it down”, line 10 in Translation G). Similarly, it performs material processes, for example patting Kurtz on his head (*gli aveva dato un buffetto sul capo*, “had flicked him on his head”, line 10 in Translation S; *gli aveva dato un colpetto sulla testa*, “had given him a small stroke on his head”, line 5 in Translation B; *gli aveva accarezzato la testa*, “had caressed his head”, line 8 in Translation M; *gli aveva dato un buffetto sulla testa*, “had flicked him on his head”, line 6 in Translation G). The reference to the beastly representation is maintained, as well as the feminine embodiment.

However, the instantiations of the personifying tendency do not always refer to the same lexical item. Often, they refer to the multiple translations of *wilderness*. Translating *wilderness* with multiple and unrelated items has an effect also on the personification. In the case of Translation S, the personification refers to five different lexical items (*deserto vegetale*, “vegetation desert”; *foresta vergine*, “virgin forest”; *immensità selvaggia*, “wild immensity”; *mondo selvaggio*, “wild world”; *terra selvaggia*, “wild land”); three in the case of Translation B (*distesa desolata*, “desolate expanse”; *giungla*, “jungle”; *terra selvaggia*, “wild land”) and Translation M (*immensità selvaggia*, “wild immensity”; *natura selvaggia*, “wild nature”; *terra selvaggia*, “wild land”); two in the case of Translation G (*landa selvaggia*, “wild land”; *selva selvaggia*, “wild woods”). This alteration potentially affects the identification of the personification as referring always to the same fictional entity, with consequences for the overall cohesion of the fictional representation. However, this happens to different extents in each of the TTs. Assumedly, the disruptive effect on the cohesion is stronger in the case of Translation S, where the lexical network is spread over five different items, while it is moderate in Translations B and M, and weak in Translation G, where *wilderness* is consistently translated with only two different items.

The effects on the personification deriving from translating *wilderness* with multiple items are clearly visible if we compare the translations of the two extracts in Examples (3) and (4) in Section 3 above. Consider their version in Translation M:

(3M) La *terra selvaggia* gli aveva accarezzato la testa ed eccolo lì, calvo come una palla – una palla d’avorio; lo aveva sfiorato e – guarda – era appassito; se l’era preso, l’aveva amato, abbracciato, gli era entrata nelle vene, aveva consumato la sua carne e sigillato la sua anima alla propria attraverso gli inimmaginabili cerimoniali di qualche iniziazione diabolica.

“The *wild land* had caressed his head and here he is, bald like a ball – an ivory ball; it had touched him lightly and – look – he had withered; it had taken him, had loved him, hugged, it had got into his veins, had consumed his flesh and sealed his soul to its own by the unimaginable ceremonies of some devilish initiation.”

(4M) E nel silenzio che era caduto improvviso su quella terra addolorata, sull’*immensità selvaggia*, sembrava che quel corpo colossale dalla vita misteriosa e feconda la guardasse, pensoso, come se stesse osservando l’immagine della propria anima tenebrosa e appassionata.

“And in the silence that had fallen suddenly on that sorrowful land, on the *wild immensity*, it seemed that that colossal body of fecund and mysterious life looked at her, pensive, as if it was observing the image of its own tenebrous and passionate soul.”

Example (3M) reproduces faithfully the personification of the African jungle as a seductive yet diabolic female entity, translating literally the lexical items that convey the trope. The extract depicts the *terra selvaggia* (“wild land”) seducing and bewitching Kurtz, luring him so as to get into his veins, to tie him to itself by *inimmaginabili cerimoniali* (“unimaginable ceremonies”). The *terra selvaggia* (“wild land”) touches Kurtz lightly, caresses him, takes him, loves him, hugs him; at the same time, it withers his scalp and consumes his flesh – it seals his soul to its own. As a result, the *terra selvaggia* (“wild land”), an inanimate entity, is ascribed with human behaviours and cognitive skills.

However, such a reference is missed later in the text, as in Example (4M) the personification has a different referent, namely *immensità selvaggia* (“wild immensity”). In this case, the direct parallelism, the embodiment of the wilderness in the figure of Kurtz’s African mistress, remains an independent instance, instead of one of a series of related personification tropes referring to *wilderness* throughout the text. Overall, the repetition of this phenomenon in various parts of the TTs affects the cohesion of the fictional representation of the African jungle as a personified entity.

4.4 Shifts and manipulation

This comparative analysis has shown that, even though the lexico-semantic patterns identified in the ST are maintained in the TTs, the fictional representation of the African jungle as a “place of darkness” and as a personified entity is less cohesive in translation. The TTs reproduce the original semantic preferences but use multiple translations for *wilderness* and *darkness*, with a consequent disruption of the link between the semantic fields and the items they refer to. As a result, the textual depiction of Africa is different from that of the original, and this may affect the way the target reader perceives the major theme “Africa and its representation”. Could this be labelled as manipulation?

In Chapter 1, I have argued that manipulation is a fuzzy concept, but scholars usually agree on a number of features that define it. For example, manipulation is said to affect mainly literary translation, as literature is more likely to be manipulated and to manipulate given its cultural and ideological implications (Dukāte 2009: 151). In this respect, the alterations identified in this analysis fit this description, as they take place in the context of literary translation. However, manipulation is also and most typically considered to be motivated and, in the majority of cases, this motivation is cultural, political, or ideological (Dukāte 2009: 46). According to the ‘manipulation school’ (cf. Chapter 1, Section 7), the translator manipulates the text in accordance with or in opposition to a cultural trend, a critical interpretation, an ideology, etc. From this point of view, the findings of this chapter do not match the definition of manipulation as ideologically motivated. On the contrary, the alterations seem to be the result of linguistic differences between English and Italian. The lack of a direct

one-to-one equivalent for *wilderness* and *darkness* forces the translators to make a choice on how to render these items. Whereas with *darkness* they have to pick an option from a range of several translation equivalents, with *wilderness* there is no equivalent at all, so the translators have to replace the term and/or paraphrase it. In translation studies literature, this phenomenon has more commonly been defined as a ‘shift’, rather than manipulation.

Shifts can be defined as deviations between the ST and the TT that can take place “at such linguistic levels as graphology, phonology, grammar, and lexis” (Venuti 2004: 148). They occur when there is a “departure from formal correspondence” (Catford 1965: 73), that is, when a TL category or item does not “occupy, as nearly as possible, the ‘same’ place in the economy of the TL as the given SL category [or item] occupies in the SL” (Catford 1965: 32). Toury (2004: 208) distinguishes between obligatory and non-obligatory shifts. Obligatory shifts are linguistically motivated, while non-obligatory shifts can be due to the translator’s literary, ideological, or cultural considerations (Dukāte 2009: 52). According to van den Broeck (2014), only optional shifts can be studied as indicators of the “translator’s preoccupation with creating an ‘acceptable’ target text, i.e. a text conforming to the norms of the target system” (van den Broeck 2014: 57), while obligatory shifts do not interfere with the adequacy of the TT.

This understanding of the difference between obligatory and non-obligatory shifts is in line with the general understanding of the difference between shift and manipulation. Obligatory shifts are not manipulation, because they are caused by linguistic constraints and therefore they are not ideologically motivated. Non-obligatory shifts can instead be the result of the translator’s choice, can be motivated and, as such, they can indicate manipulation. However, Dukāte (2009) argues that manipulation can also be unconscious and not intentionally motivated by the translator, for example as a result of “the workings of human psyche” or of errors and mistakes (Dukāte 2009: 87). Linguistic and factual mistakes or ignorance on the part of the translator, although unintentional, can have manipulative effects. Dukāte (2009: 115) names this type of manipulation ‘text-internal translational manipulation as unconscious distortion’ and defines it as a “manipulation due to the translator’s lack of professionalism, and is manifested as errors, which seriously mislead the reader and distort the original text”.

The choice of translating *wilderness* and *darkness* with multiple terms, instead of using the same item, can be considered an error on part of the translators: they failed to recognise the importance of the reiteration, the function of the lexico-semantic patterns as a key device to make the fictional representation of the African jungle cohesive. Given the absence of an ideological motivation, this alteration could be easily dismissed as an example of linguistic shift. However, in line with Dukāte's (2009) definition of manipulation as unconscious distortion, I argue that this shift produces manipulative effects. If these alterations are seen in terms of the effect they produce on the target readership, as opposed to the reason why they were introduced in the first place, then it does not make any difference whether they are motivated or simply due to misrecognition. In both cases, they produce a distortion on the original text, from the point of view of the target reader. Without realising it, the translators modified a formal feature that is linked to a major thematic concern in HoD. As a consequence, the way the reader perceives this theme can be affected. Whether the translators make this by mistake or by following a personal agenda, the effect would not be different for the target reader. In fact, as Dukāte (2009: 88) explains, "an error or several errors can have a cumulative manipulative effect and as a result the text may seem to be factually, linguistically or ideologically manipulated".

I would argue that the way manipulation has been generally regarded neglects the manipulative effects that even unconscious and/or unmotivated linguistic shifts can have. The outcome of my analysis suggests that manipulation can occur more elusively than is usually thought. Manipulative effects like the one discussed here may even be more common than what is usually defined as manipulation in translation studies research. This is simply because linguistic shifts occur in all sorts of translations, at different levels. In Chapter 1, I have raised the question whether manipulation mainly occurs in contexts in which the balance of power between source culture and target culture is so uneven that a manipulative phenomenon is almost unavoidable (see Munday 2008 or Tymoczko 2010), as the literature on the topic seems to imply. My findings seem to challenge this view, although further research is needed to answer this question. However, the outcome of my analysis suggests that manipulative effects resulting from linguistic shifts can play an equally fundamental role in altering the TTs as ideological and/or political biases. Finally, the use of corpus methodologies to study manipulation has allowed me to gain a linguistic close-

up view of the phenomenon, showing concretely how lexical alterations may have manipulative effects, even when they are just the result of mere mistakes or misrecognitions.

5. Conclusion

This chapter has demonstrated that the items resulting from the KW analysis and categorisation in Chapter 4 can be successfully analysed to establish a link between textual patterns and the fictional world. The study of the “Africa words” has shown how their use in HoD is interconnected to construct a cohesive fictional representation of the African jungle. The four words share the same semantic preferences and co-occur with related vocabulary: I have argued that these lexico-semantic patterns (and the local textual function they jointly enact) are central to the interpretation of the major theme “Africa and its representation”. The chapter has linked quantitative and qualitative analyses, showing both how the patterns develop throughout the entire text and how they are instantiated in individual textual extracts. This double perspective has enhanced our understanding of the role that the lexical level of HoD plays in building up the fictional world and triggering thematic concerns. I have provided a detailed account of the precise stylistic tools that instantiate the fictional representation of the jungle as “a place of darkness” and as a personified entity. These new insights into Conrad’s writing have been backed by the discussion of more traditional literary criticism.

The comparison of the TTs has shown the effects of translation on the representation of the jungle. The analysis has indicated that all four TTs maintain the dominant semantic fields identified in the original, probably as a result of a literal translation of Conrad’s lexical choices. However, I have argued that translating *wilderness* and *darkness* with multiple items disrupt the lexical network that the original establishes, with a consequence on the overall cohesion of the TTs. In translation, the fictional representation of the African jungle is less cohesive than in the original; therefore, the jungle – its connection with *wilderness* and *darkness* – may be perceived less as a unitary fictional entity. The comparison has also suggested that shifts and misrecognitions on the part of the translators can have manipulative effects.

The corpus-assisted approach has proved to be useful in pointing out this outcome, as it has enabled me to look closely at the relation existing between alterations at the linguistic level and manipulations at the interpretative level.

CHAPTER SIX

“Black things”: The fictional representation of the African natives

1. Introduction

In this chapter, I will examine how the African natives are linguistically represented in HoD and in its four Italian translations. The investigation will be based on the outcome of the KW analysis (Chapter 4) and it will concentrate on the “African words” (*nigger, niggers, negro, savage, savages, native, natives, black, and blacks*).

This chapter is divided into two main parts. In the first part (Section 2), I will focus on the analysis of HoD. I will start, in Section 2.1, by investigating the usage of the “African words” in the ST and compare their frequency in the reference corpora. I will aim to test whether the frequency of the “African words” in HoD is statistically deviant (Leech & Short 2007) compared to their frequency in CC and RC. In Section 2.2, I will carry out a concordance analysis in order to identify the semantic preferences and prosodies of the “African words”. I will show that the “African words” and their lexico-semantic patterns play a central role in defining how the natives are represented in HoD. Specifically, I will argue that these lexico-semantic patterns build up a dehumanising fictional representation of the Africans (Section 2.3).

In the second part of the chapter (Section 3), I will focus on the analysis of the four Italian translations. To start with, in Section 3.1, I will identify, compare, and discuss the translations of the “African words”. In Sections 3.1.1 to 3.1.5, I will concentrate on each “African word” across the TTs, examining whether the translators’ lexical choices produce a shift in the TT. In particular, I will focus on those shifts that can intensify or mitigate the potential racial implications of the ST word. In Section 3.2, I will discuss the general tendencies resulting from the individual yet repeated shifts, analysing their effect of the overall fictional representation of the Africans. I will then investigate whether the TTs reproduce equivalent semantic preferences and prosodies as those identified in the ST, carrying out a concordance analysis of the Italian “African words” (Section 3.3). Finally, in Section 4, I will close the chapter with a discussion of the findings.

2. The representation of the Africans in *Heart of Darkness*

In Chapter 2, I have explained that one of the most debated themes in HoD today is “race and racism”. I have argued that the way the Africans are depicted in the text is central to the critical discussion of this theme. In this section, I will analyse the representation of the natives, through the examination of the “African words”. First, I will focus on their specific use in HoD and the functions they enact in context. Then, I will show how the linguistic construction of the Africans is related to the theme of race and racism, discussing the findings of the linguistic investigation in light of the literary debate.

2.1 Frequency comparison of the “African words”

The analysis of the fictional representation of the Africans starts with a frequency comparison in order to verify whether there is a deviance in the use of the “African words” compared to their frequency in the reference corpora. Table 6.1 shows both raw and normalised frequencies in HoD, CC, and RC. A log-likelihood test⁸ is used to check whether the differences in frequency are statistically significant (p -value threshold = 0.0001). When statistically significant, the respective cell in the table is shaded in grey.

⁸ The log-likelihood calculator is available at <http://ucrel.lancs.ac.uk/llwizard.html> (last accessed August 2015).

Table 6.1 Frequency of the “African words” in HoD and in the reference corpora

	HoD		CC		LL HoD vs CC ($p < 0.0001$)		RC		LL HoD vs RC ($p < 0.0001$)	
	Raw	Norm.*	Raw	Norm.		Raw	Norm.			
nigger	5	0.13	43	0.02		12	0			
niggers	5	0.13	17	0.01		3	0			
Total	10	0.26	60	0.03	21.63	15	0		60.52	
savage	13	0.33	152	0.09		145	0.03			
savages	9	0.23	35	0.02		41	0.01			
Total	22	0.56	187	0.11	35.68	186	0.04		69.55	
negro	3	0.08	25	0.01	4.96	11	0		13.82	
black	42	1.07	1,188	0.68		1,673	0.39			
blacks	1	0.02	5	0		5	0			
Total	43	1.10	1,193	0.68	8.08	1,678	0.40		32.36	
native	2	0.05	140	0.08		199	0.05			
natives	5	0.13	43	0.02		49	0.01			
Total	7	0.18	183	0.10	1.63	248	0.06		6.14	

* Normalisation per 1,000 words

The frequencies in Table 6.1 refer to the total occurrences of the “African words” in HoD and in the reference corpora, as opposed to the occurrences of the “African words” referring specifically to the natives. This distinction concerns mainly *black*, as this word has a wider range of meanings compared to the other “African words” and it is used in ways other than referring to black people. *Black* is in fact the most frequent item among the “African words”, occurring 42 times in HoD, but it is used to refer to the Africans in 22 cases out of 42. With regard to the other “African words”, the distinction has minimal effect. *Savage* and *native* can be used as adjectives and can refer to something/someone other than the Africans, yet this only occurs a few times in HoD (see Section 2.2). *Blacks*, *savages*, and *natives* are nouns and as such their range of use is restricted compared to the adjectival forms: they always refer to the Africans, except one occurrence of *savages*. Finally, *nigger*, *niggers*, and *negro* are relevant in all cases because they always refer to the Africans.

The reason for not making this distinction at this stage is that it would be very difficult and time-consuming to disambiguate the occurrences of *black* in the reference corpora. Even with the help of a POS tagger that distinguishes between nouns and adjectives, the results would still need to be checked manually in order to identify the occurrences that refer specifically to black people. This would involve

manually analysing a very large amount of concordance lines, given that *black* occurs more than 1,100 and 1,600 times in CC and RC respectively.

With this in mind, Table 6.1 shows that all the normalised values are higher in HoD than in the reference corpora, although not all the differences are statistically significant. Starting from *nigger/niggers*, their frequency difference is significant: these items are used about 8 times more frequently in HoD than in CC and over 80 times more frequently than in RC. Similarly, a significant difference exists for *savage/savages*, used considerably more in HoD compared to both CC and RC. In contrast, the difference in the frequency of *black/blacks* is significant only when compared with that in the RC, yet not when compared to CC. However, this result may be affected by the fact that the frequency count does not distinguish between *black/blacks* referring to black people and other uses of the words. Finally, there is no statistical significance in the frequency difference of *negro* and in the frequency difference of *native/natives*. In light of this, it can be said that a deviance exists in the use of *nigger/niggers* and *savage/savages* in HoD, compared to the two reference corpora, while *native/natives* and *negro* occur in line with their expected frequency.

2.2 The “African words” in context

Assigning prominence or literary relevance to a statistical deviance involves the identification of an “artistically coherent pattern of choice” aimed at a “particular literary end” (Leech & Short 2007: 40). This process goes beyond the identification of the deviance and requires the study of the deviant feature in its context. In this section, I will carry out a concordance analysis of the “African words”, looking for the patterns they establish and the functions these patterns enact. Moreover, the concordance analysis allows me to identify and remove those occurrences that do not refer specifically to the Africans. Table 6.2 below shows the outcome of this disambiguation, reporting the frequencies of only those cases in which the words do refer to the natives. This procedure has also enabled a further level of discrimination in the analysis of the “African words”. When used as an adjective, these terms can refer either to the Africans or to someone/something else. When referring to the Africans, these adjectives are still relevant for this study, while when referring to

someone/something else they are not. It is therefore important to account for the former and exclude the latter. The concordance analysis made this distinction clear, showing when the “African words” are used directly to refer to the natives and when they are instead used to refer to them indirectly, i.e. as an attribute. For instance, in Example (1) *savage* is directly used as a noun to refer to a native:

- (1) And between whiles I had to look after the *savage* who was fireman. He was an improved specimen; he could fire up a vertical boiler.

Whereas, in Example (2) *black* is used as an adjective premodifying a noun that is not included in the “African words” (*figure*). Although indirectly, *black* still refers to an African, as an attribute that specifies his/her physical appearance.

- (2) A *black* figure stood up, strode on long *black* legs, waving long *black* arms, across the glow.

The result of this analysis is summarised in Table 6.2 below.

Table 6.2 Frequency of the “African words” in HoD referring to the natives

“African words”	Nouns	Attributes
nigger	5	0
niggers	5	0
savage	3	4
savages	8	0
negro	3	0
black	2	20
blacks	1	0
native	1	0
natives	5	0

Table 6.2 shows that, in the vast majority of cases, the “African words” do refer to the natives, especially when they are used as nouns. *Savage* is mostly used as an attribute for the natives, while in 6 cases it does not refer to the natives at all (cf. Table 6.1). The only item whose frequency displays a clear drop is, as expected, *black*. There are only 2 instances where it is used as a noun for black person, while the remaining 20

occurrences are indirect references. Taking into consideration the use of the “African words” as nouns to indicate directly a native (that is, excluding the attribute uses), a tendency can be identified: the African natives that populate the book’s pages are more frequently called *nigger/niggers*, *negro*, or *savage/savages* than *black/blacks* or *native/natives*.

With the non-relevant uses of the “African words” removed, the remaining instances are examined in order to delineate their semantic and pragmatic profile. Similarly to what I have shown with the “Africa words” (cf. Chapter 5), the concordance analysis of the “African words” highlights the presence of four repeated semantic preferences, shared by all the items in the group. I call the first semantic field “physicality”; see Figure 6.1 below.

1	in the station was in a muddle--heads, things, buildings. Strings of dusty niggers with splay feet arrived and departed; a stream of manufactured
2	say I saw any road or any upkeep, unless the body of a middle-aged negro , with a bullet-hole in the forehead, upon which I absolutely stumbled
3	civil nor uncivil. He was quiet. He allowed his 'boy'--an overfed young negro from the coast--to treat the white men, under his very eyes, with
4	, but nothing to fasten them with. And every week the messenger, a long negro , letter-bag on shoulder and staff in hand, left our station for the
5	and left to the impressed pilgrims. A quarrelsome band of footsore sulky niggers trod on the heels of the donkey; a lot of tents, camp-stools, tin
6	the matter to their satisfaction. Their headman, a young, broad-chested black , severely draped in dark-blue fringed cloths, with fierce nostrils and

Figure 6.1 “Physicality” semantic field

The “physicality” semantic field includes body part words, such as *feet* (line 1), *body* (line 2), and *shoulder* (line 4), and also items referring more generally to the bodily appearance of the natives. In the sample of concordance lines in Figure 6.1, these are represented by *overfed* (line 3), *footsore* (line 5), and *broad-chested* (line 6).

The second semantic field is labelled “Collectives” as it includes collective nouns used to refer to the African natives as a homogeneous group, as opposed to different individuals. This is illustrated below in Figure 6.2 with the words *crowd* (lines 2 and 6), *band* (line 3), *strings* (line 4), and *lot* (line 5).

2	at last of asserting his self-respect in some way. Therefore he whacked the old nigger mercilessly, while a big crowd of his people watched him, thunderstruck, till
3	right and left to the impressed pilgrims. A quarrelsome band of footsore sulky niggers trod on the heels of the donkey; a lot of tents, camp-stools, tin boxes,
4	else in the station was in a muddle--heads, things, buildings. Strings of dusty niggers with splay feet arrived and departed; a stream of manufactured goods,
5	a hut. The population had cleared out a long time ago. Well, if a lot of mysterious niggers armed with all kinds of fearful weapons suddenly took to travelling on the
6	staggered forward again, and almost at the same time I noticed that the crowd of savages was vanishing without any perceptible movement of retreat, as if the

Figure 6.2 “Collectives” semantic field

The third semantic field has been called “incomprehensibility”. This field collects words referring to the sounds produced by the natives, perceived by the non-natives as non-language and incomprehensible. For example, in Figure 6.3 below, this field is represented by *groaned* (line 1), *moaned* (line 2), *yells* (line 3), and *howling* (line 6). This field also includes terms that indicate directly this lack of comprehension between the colonisers and the natives, for instance, *incomprehensible* in line 5.

2	glow, whence proceeded a sound of hissing; steam ascended in the moonlight, the beaten nigger groaned somewhere. 'What a row the brute makes!' said the indefatigable man with
3	very heart-its mystery, its greatness, the amazing reality of its concealed life. The hurt nigger moaned feebly somewhere near by, and then fetched a deep sigh that made me
4	there would be a glimpse of rush walls, of peaked grass-roofs, a burst of yells, a whirl of black limbs, a mass of hands clapping of feet stamping, of bodies swaying, of eyes rolling,
5	droop of heavy and motionless foliage. The steamer toiled along slowly on the edge of a black and incomprehensible frenzy. The prehistoric man was cursing us, praying to us,
6	extravagance of emotion, even such as I had noticed in the howling sorrow of these savages in the bush. I couldn't have felt more of lonely desolation somehow, had I been

Figure 6.3 “Incomprehensibility” semantic field

Finally, the last and larger semantic field, “general negative” (Figure 6.4), gathers a wide variety of words that have negative implications. It includes terms referring to specific negative actions (*whacked* and *beaten*, lines 8 and 6 respectively), negative descriptions (*unhappy* and *fool*, lines 7 and 5 respectively), and also words indicating negative states or feelings (*indifference* and *sorrow*, lines 7 and 1 respectively).

4	of emotion, even such as I had noticed in the howling sorrow of these savages in the bush. I couldn't have felt more of lonely desolation somehow
5	was yet full of noise and smoke when I made a dash at the wheel. The fool-nigger had dropped everything, to throw the shutter open and let off
6	a sound of hissing; steam ascended in the moonlight, the beaten nigger groaned somewhere. 'What a row the brute makes!' said the
7	, without a glance, with that complete, deathlike indifference of unhappy savages. Behind this raw matter one of the reclaimed, the product of the
8	of asserting his self-respect in some way. Therefore he whacked the old nigger mercilessly, while a big crowd of his people watched him,

Figure 6.4 “General negative” semantic field

Table 6.3 below collects all the collocates of *nigger/niggers*, *savage/savages*, *negro*, *native/natives*, and *black/blacks* in a span of 5:5 that form part of the “physicality”, “collectives”, “incomprehensibility”, and “general negative” semantic fields.

Table 6.3 Collocates of the “African words” belonging to the four semantic fields

Physicality
feet, footsore, old, body, middle-aged, overfed, young, long, shoulder, head, broad-chested, athletic, neck, necked, limbs, hands, legs, arms, eyelids, wild-eyed, face, bones.
Collectives
crowd (2)*, strings, lot, band, camp, picket, tribe, file.
Incomprehensibility
howling, mysterious, groaned, row, moaned, yells, incomprehensible, frenzy.
General negative
beaten (2), hate (2), suppression (2), unrestrained, grief, sorrow, deathlike, indifference, unhappy, death, dusty, quarrelsome, sulky, hurt, fool, whacked, mercilessly, enemies, scuffle, insolent, disease, listlessly, starvation.

* Frequency, when ≥ 2 , in brackets

The fictional representation that these fields instantiate is strong enough throughout the text that it can be easily identified in contexts longer than the five words to the left and to the right of the core taken into account with the concordance analysis. However, I will address this point in more detail in Section 2.3, as I want to highlight another pattern emerging from the study of *black*, which will be relevant to the interpretation of the fictional representation of the natives in the next section. As indicated in Table 6.2, *black* is used only twice as a noun to refer to a black person, while in the majority of cases (20 occurrences) it refers to the natives indirectly, as an attribute. Specifically, *black* often occurs in the construction “*black*[ADJ] + common noun”, such as *black figure* in Example (2) above. Figure 6.5 below shows all 20 occurrences of *black* as an attribute.

It can be noted that *black* does not co-occur only with words that are usually employed to indicate men or women, as in *black men* (line 1), *black fellows* (lines 2, 8, and 12), and *black people* (lines 4 and 9). On the contrary, it occurs more frequently with abstract terms, for example *figure/figures* (lines 3 and 14), or even with words which can hardly be considered synonyms of *men*, *fellows*, or *people*: *whirl of black limbs* (line 5), *black things* (line 7), *black bones* (line 10), *black shadows* (line 11), and *black shapes* (line 13). This pattern is identified by Kujawska-Lis (2008) as well, although her study does not provide any quantification to support the claim. Kujawska-Lis (2008: 168) hypothesises that Conrad might have preferred to use a vocabulary that emphasised clearly the belonging to the native race, such as

nigger and *negro*, limiting the use of *black* to “metaphorical” meanings, as in the constructions showed above.

1	"A slight clinking behind me made me turn my head. Six blackADJ men advanced in a file, toiling up the path.
2	the contrast of expressions of the white men and of the blackADJ fellows of our crew, who were as much
3	out suddenly, and we went outside. The moon had risen. BlackADJ figures strolled about listlessly, pouring water
4	scene of inhabited devastation. A lot of people, mostly blackADJ and naked, moved about like ants. A jetty
5	walls, of peaked grass-roofs, a burst of yells, a whirl of blackADJ limbs, a mass of hands clapping of feet
6	deliberately to the first I had seen--and there it was, blackADJ , dried, sunken, with closed eyelids--a head
7	blue space, sparkling with dew and starlight, in which blackADJ things stood very still. I thought I could see a
8	discretion with great gravity. 'I have a canoe and three blackADJ fellows waiting not very far. I am off. Could you
9	path was steep. A horn tooted to the right, and I saw the blackADJ people run. A heavy and dull detonation shook
10	Then, glancing down, I saw a face near my hand. The blackADJ bones reclined at full length with one shoulder
11	not criminals, they were nothing earthly now--nothing but blackADJ shadows of disease and starvation, lying
12	one a momentary contact with reality. It was paddled by blackADJ fellows. You could see from afar the white of
13	of the launched earth had suddenly become audible. " BlackADJ shapes crouched, lay, sat between the trees
14	We were within thirty yards from the nearest fire. A blackADJ figure stood up, strode on long blackADJ legs,
15	foliage. The steamer toiled along slowly on the edge of a blackADJ and incomprehensible frenzy. The prehistoric
16	stood up, strode on long blackADJ legs, waving long blackADJ arms, across the glow. It had horns--antelope
17	nearest fire. A blackADJ figure stood up, strode on long blackADJ legs, waving long blackADJ arms, across the
18	and faces. Suddenly the manager's boy put his insolent blackADJ head in the doorway, and said in a tone of
19	idea at all connected with it? It looked startling round his blackADJ neck, this bit of white thread from beyond the
20	not hear, he frowned heavily, and that frown gave to his blackADJ death-mask an inconceivably sombre, brooding,

Figure 6.5 Concordance lines of *black* used as an adjective

Overall, the textual behaviour of the “African words” is, in some respects, similar to that of the “Africa words” (cf. Chapter 5, Section 2.3). In this case too, every time the reader comes across *nigger/niggers*, *savage/savages*, *negro*, *native/natives*, and *black/blacks*, he/she will always find them within the same semantic fields in their contexts of occurrence. This repeated lexico-semantic pattern, when maintained throughout the whole text, establishes a persistent fictional representation of the African natives. The depiction of the Africans emphasises their body and bodily appearance, rather than what they think or feel; there is a tendency to refer to them as a homogeneous and undistinguished group, rather than individuals; they are represented as producing unintelligible noises impossible to comprehend; lastly, they are associated with negative and unpleasant situations or states of mind. These semantic preferences, and the resulting negative semantic prosody, build a dehumanising fictional representation of the African natives, which closely mirrors Achebe’s (1990) interpretation.

2.3 Dehumanisation in the fictional representation of the natives

In Chapter 2, I have argued that one of the key points on which Achebe's (1990) criticism hinges is the way African natives are portrayed in *HoD*. The passages in which they appear are, according to this critic, the "most revealing" (Achebe 1990: 5) of Conrad's racist attitude. Natives' descriptions emphasise the physicality of their appearance, and never reveal their inner thoughts or psychology; in fact, descriptions in which they are not "just limbs or rolling eyes" are rare (Achebe 1990: 6). They are deprived of human expression and language altogether (Achebe 1990: 8), babbling and grunting frenziedly even among themselves. They disappear, from the general picture, as human factor and are reduced to the role of props (Achebe 1990: 12). In the final analysis, Achebe (1990: 12) claims that Africans are dehumanised in the text and that *HoD* is a novel that celebrates this dehumanisation.

The question of dehumanisation has been touched upon by many other scholars, who agree and disagree on this issue similarly to the way critics have agreed and disagreed with Achebe's (1990) overall argument. For example, prior to Achebe's (1990) public lecture, Curle (1914) praises Conrad's skills in representing non-European people as neither dehumanised nor Europeanised (Curle 1914: 121, see Chapter 2 for more details). Hawkins (2006), in contrast, agrees with Achebe (1990) and recognises dehumanisation in *HoD*. The dehumanising tendencies he identifies (Hawkins 2006: 366) are also very similar to those pointed out by Achebe (1990): none of the natives have a proper name; none of them are given more than one paragraph (apart from Kurtz's assumed African mistress); there is no representation of their thoughts or points of view, partly because they never speak, apart from four pidgin sentences in the whole text. Lawtoo (2012) also identifies the presence of dehumanisation in the representation of the natives, although the way he points it out is less straightforward. He maintains that the text oscillates continuously back and forth "between racist injunctions that dehumanize racial others and repeated attempts to nuance such racist distinctions" (Lawtoo 2012: 248).

The semantic preferences discussed in the previous section seem to instantiate linguistically that dehumanising process Achebe (1990), Hawkins (2006), and Lawtoo (2012) recognise. What is more, the semantic fields identified match quite accurately what the critics have referred to when breaking down the dehumanisation into specific

tendencies or attitudes. The “physicality” semantic field reflects that representative focus on the Africans’ bodily appearance and materiality as opposed to their consciousness. An instance of this tendency has been provided in Example (2) above (included also in Figure 6.5, line 14), where the description returns insistently to the native’s black limbs (*black legs, black arms*), stressing his physical appearance. As I have argued in the previous section, this fictional representation is enacted so pervasively throughout the text that it is possible to identify further instantiations also in extracts other than those taken into account with the concordance analysis. Instances of dehumanisation (or the emphasis on physical description, as in this case) are recognisable also where the natives are referred to in terms other than the “African words” or in portions of the text longer than the 5:5 span considered for the creation of the semantic fields. Consider Example (3) below.

- (3) Then, glancing down, I saw a *face* near my hand. The *black bones* reclined at full length with one *shoulder* against the tree, and slowly the *eyelids* rose and the sunken *eyes* looked up at me, enormous and vacant, a kind of blind, white flicker in the depths of the *orbs*, which died out slowly.

For the identification of the “physicality”, “collectives”, “incomprehensibility”, and “general negative” semantic fields, a span of 5:5 has been adopted, that is, only words that occurred within the limit of five words to the left and five words to the right of the core have been counted. In Example (3), only *face* and *bones* have been included in the “physicality” semantic field, as they occur in a span of 5:5 from *black*. However, the stress on the bodily description is not limited to these two items. Extending the analysis to the whole passage, it can be noted that all of the references to the African are related to a part of his body – besides *face* and *bones*, there is also *shoulder, eyelids, eyes, and orbs*. This extract also evokes the familiar metaphor associated with the eyes, “the eyes are the mirror of the soul”, in order to subvert it and to intensify the focus on the body. According to this metaphor, the eyes provide access beyond the exterior appearance of an individual to the inner self and spirit. In Example (3), Marlow looks at the African’s face but, when the eyelids slowly rise, he

only sees an enormous, blinding void, as if there is nothing to grasp inside that bodily shell, as if there is no soul inside to be mirrored in the orbs of that figure.

In Vermeule's (2010) definition of 'mind blindness' – one of the commonest tropes of dehumanisation – the deprivation of mind takes centre stage: "the point is to deny other people the perspective of rational agency by turning them into animals, machines, or anything without a mind" (Vermeule 2010: 195). The emphasis on the "limbs or rolling eyes" (Achebe 1990: 6) subordinates the mind of the natives to their physical appearance, or even denies its presence altogether. This effect is corroborated by the natives' lack of language, as pointed out by Achebe (1990: 8) and Hawkins (2006: 366), and as reflected by the "incomprehensibility" semantic field. The present analysis shows that the Africans, rather than speaking, moan (Figure 6.3, line 3), yell (line 4), frenzy (line 5), or howl (line 6). Depriving the natives of language not only undermines their cognitive ability, but questions their very humanity by implying that they are closer to howling and frenzied beasts than people.

Again, this tendency is easily recognisable also in extended contexts, within passages that have not been looked at through the concordance analysis, because no "African word" appears in them. For instance, Example (4) below is taken from the final part of the text: Marlow, after having taken Kurtz on board, is sailing out of the station, while a group of natives stares puzzled at the steamer.

- (4a) When we came abreast again, they faced the river, stamped their feet, nodded their horned heads, swayed their scarlet bodies; [...] they *shouted* periodically together strings of *amazing words* that resembled *no sounds of human language*; and the deep *murmurs* of the crowd, interrupted suddenly, were like the responses of some *satanic litany*.

Followed by:

- (4b) There was an eddy in the mass of human bodies, and the woman with helmeted head and tawny cheeks rushed out to the very brink of the stream. She put out her hands, *shouted* something, and all that wild mob took up the *shout* in a *roaring chorus* of *articulated, rapid, breathless utterance*.
- “Do you understand this?” I asked.

“Do you understand this?” (Conrad [1902] 1994: 96), a bewildered Marlow asks Kurtz, after witnessing what seems to him an incomprehensible frenzy. The way Marlow describes it in Example (4a) makes it clear: these shouts resemble *no sounds of human language*. In just a few lines, these passages gather many items that would fit the “incomprehensibility” semantic field very well. The natives *shouted* and produced *murmurs*, as well as *a roaring chorus* of *amazing words* and *breathless utterances* that for Marlow and his crew sounds almost like a *satanic litany*. This proves further that the tendencies unveiled by the concordance analysis are used pervasively throughout the whole text, and not just in the short contexts of the concordance lines. What is more, these extracts show how the various aspects of the fictional representation, the four semantic fields, interact and co-occur. The “physicality” semantic field and its focus on the bodily appearance are also clearly notable in Example (4). The fictional representation of the natives stresses again their physical exteriority – their *bodies*, *stamping feet*, and *horned heads*; in the case of the woman, her *head*, *cheeks*, and *hands*.

The extracts in Example (4) show also another one of the tendencies signalled by the critics as contributing to the dehumanization of the natives, and similarly represented by one of the semantic fields identified: the tendency to refer to the Africans as a homogeneous group instead of individuals, instantiated by the “collectives” semantic field. The natives are referred to as *crowd* in Example (4a), and *mass* and *wild mob* in Example (4b). Hawkins’s (2006: 366) observation that no native is referred to with a proper noun is related to this, as proper nouns signal individuality and uniqueness; on the contrary, this uniqueness is lost in a collective noun, which merges together differences and individualities until they are homogeneously flat. Example (5) below shows another instance of this process in a

passage not accounted for in the concordance analysis as, again, no “African word” occurs in it.

- (5) When next day we left at noon, the *crowd*, of whose presence behind the curtain of trees I had been acutely conscious all the time, flowed out of the woods again, filled the clearing, covered the slope with a *mass* of naked, breathing, quivering, bronze bodies.

Crowd and *mass* are the collective nouns used in this extract to refer to the natives. The image conveyed in Example (5) is that of a homogeneous fluid that pours out of the forest, with lexical choices such as *flowed*, *filled*, and *covered* enacting this image. This indistinct and indistinguishable *crowd* fills the slope with *a mass of naked, breathing, quivering, bronze bodies*, recalling again the emphasis on the natives’ physicality.

Finally, the “general negative” semantic field does not directly represent any of the observations raised by the critics. There is no specific tendency or aspect of the dehumanisation that mirrors it, in contrast to the other three fields that instead seem to be direct linguistic instantiations of the dehumanising attitudes identified by Achebe (1990) and Hawkins (2006). Nevertheless, the “general negative” field acts as the adhesive for the other tendencies. It is the largest field and is widely spread across the whole text, sometimes occurring simultaneously with the other fields. This strengthens the negative semantic prosody that the “African words” – or the overall fictional representation of the Africans in general – possess.

According to Vermeule (2010), mind blindness, or dehumanisation in general, is a tool of emotional dominance, common in the context of war or oppression where opposing groups of people demonise each other. Moral conventions and social behaviours are valid and shared within one group, but tend to be overlooked or totally negated to the opposing group. When dehumanisation takes place, “the members of the hated counter group do not count as human, and therefore moral rules do not apply to them” (Vermeule 2010: 195). This could happen for many reasons: from the intellectual stimulation this process might encompass to the need of showing what an individual is really like by comparing it to something other than human. However,

Vermeule (2010: 200) asserts that the most important reason why dehumanisation happens is to “assert mastery over the other person”. In fact, shifting the ontological status of a person from a civilised and rational human being to a brute, animal, or machine facilitates that process of moral justification necessary to perpetrate domination. This is especially true in the context of colonial domination, where the motif of the “civilising mission” had been the constant hypocrisy which has legitimated seeing “those who have a different complexion or slightly flatter noses than ourselves” (Conrad [1902] 1994: 10) as “sullen peoples, half devil and half child” (Kipling [1899] 2013: 111) in need of guidance and authority.

In the light of this, the dehumanising representation of the African natives in HoD can be seen as a reflection of that “emotional dominance” that the colonisers impose on the colonised. Its linguistic instantiation, i.e. the semantic fields, echoes that socio-culturally established and generally accepted vision of the non-European other as intellectually inferior and emotionally unstable, driven by instinct rather than reason and underdeveloped on a scale of assumed civilisation.

3. The comparison of the translations

In this section, I will compare the TTs on the basis of the findings of the ST analysis. First, in Section 3.1, I will study the translators’ lexical choices to translate each of the “African words”. Second, in Section 3.2, I will consider all the lexical choices together in order to analyse whether they constitute an overall tendency in the TTs. Finally, in Section 3.3, I will examine the semantic preferences and prosody of the Italian “African words”, comparing their function with the function they enact in the original. The overarching aim of Section 3 is to study the consequences of the TT textual alterations on the fictional representation of the African natives, and how these can affect the theme “race and racism” in HoD.

The comparison between the ST and the TTs revolves around the identification of shifts in the translators’ lexical choices, i.e. whether the translators opted for a word other than the closest equivalent to translate the “African words”. For example, the direct equivalent of *nigger* in Italian is *negro*. A shift would occur if a word other than *negro* (“nigger”) is used to translate *nigger*, like *nero* (“black”).

When repeated, shifts like this can be indicative of manipulative phenomena, therefore they are relevant to look at to study translation manipulation.

This approach based on the analysis of shifts has been followed by other studies with similar aims to mine. For example, in Chapter 4 (Section 5.2), I have discussed a study by Kujawska-Lis (2008: 167) in which she looks at shifts in the translation of HoD into Polish. She focuses on terms that can be related to the accusation of racism: *black*, *savage*, *nigger*, *negro*, and *brute*. She compares two Polish translations to test whether Achebe's (1990) criticisms have affected the translators' lexical choices. Kujawska-Lis (2008: 169) shows that, in the early translation, *nigger* is translated with the Polish *Murzyn*, which simply describes a person of black African race. However, in the modern translation, *nigger* is translated with its Polish equivalent, *czarnuch*. She argues that, in the case of the early translation, a neutral TL term (*Murzyn*) is used to replace an abusive ST item (*nigger*), while in the case of the modern translation, the abusive word is translated with its similarly abusive Polish equivalent (*czarnuch*): these choices are seen as having the potential to affect the perception of the text as racist at the lexical level (Kujawska-Lis 2008: 169).

A similar method is adopted by Trupej (2012) in a comparative study of the translation of *nigger* in two Slovenian versions of Mark Twain's *Adventures of Huckleberry Finn*. Trupej (2012: 93) examines shifts in the translation of *nigger* and discusses the effects they have on the construction of characters and plot in the TTs. The 1948 translation displays numerous instances of translational shifts, as the terms chosen to translate *nigger* intensify the racist discourse (Trupej 2012: 98). At times, *nigger* is replaced with items such as *prekleti črni ubijalec* ("damned black killer"), *smrdljiva črna drhal* ("stinking black rabble"), or *črni malopridneži* ("black good-for-nothing"). In other cases, the personal pronoun *he* is replaced by *zamorec*, the Slovenian equivalent of *nigger* (Trupej 2012: 98). Trupej (2012: 99) defines these shifts "explicit intensifications of racist discourse" and argues that, as a consequence of them, both positive and negative characters appear more racist in translation. On the contrary, in the 1962 TT, *nigger* is often omitted, replaced by a personal pronoun, or by *črnc*, the Slovenian neutral term for a black person (Trupej 2012: 101). Trupej (2012: 102) considers these shifts as "less severe" lexical choices that "significantly reduce the severity of the racism displayed in the novel".

Mouka et al. (2015) employ corpus tools to explore shifts in the context of racist discourse in film translation. Except for their use of corpora, Mouka et al.'s (2015) approach is similar to Kujawska-Lis's (2008) and Trupej's (2012): they compare the ST to the TTs in order to identify shifts, which are then classified as neutral transfer, over-toning, or under-toning depending on the translators' lexical choices (Mouka et al. 2015). Corpus tools are used to retrieve the examples of racist discourse from the films analysed, while reference corpora are employed to help interpreting how racist markers are used in the cross-linguistic contexts, but only "as far as possible" (Mouka et al. 2015: 48). In fact, they acknowledge that establishing the strength of each instance of racist discourse is "the most subjective parameter" of their method (Mouka et al. 2015: 56), even with the help of reference corpora. Their analysis focuses on five films (*Do the Right Thing*, *American History X*, *Monster's Ball*, *Crash*, and *This is England*) and their Spanish and Greek translations. For some of the shifts discussed, they too look at the word *nigger*. For example, they look at a passage from *American History X*, *We'll let the niggers, kikes and spics grab for their piece of the pie*, and its Spanish translation, *Dejemos que negros y latinos se lleven su parte* ("Let the blacks and Latinos get their part"). They mark the use of *nigger* in the original as "highly negative" (Mouka et al. 2015: 60), while consider *negros* ("blacks") a shift that neutralises the derogatory connotation of the original utterance. Evaluating all the shifts in a similar fashion, they conclude that the general tendency in their data is to mitigate or omit racial slurs, although some instances of over-toning are identified as well (Mouka et al. 2015: 63).

In line with these studies, I will compare HoD to the four TTs to identify shifts in the translation of the "African words". When a shift is recognised, I will assess whether it intensifies or mitigates the potential racist implications of the original ST item. In order to help to do so as far as possible, I will use corpus evidence from an Italian reference corpus, *itTenTen* (Jakubiček et al. 2013). *itTenTen* is a large Italian corpus (3 billion words) built with texts collected from across the internet, inclusive of a wide range of text types and registers. *itTenTen* is available as a pre-loaded corpus on the *Sketch Engine* platform (Kilgarriff et al. 2014).

3.1 The “African words” in translation

To start with, I compare the ST and the TTs in order to identify the translations of the “African words”. Only the items referring to the Africans have been taken into consideration (those counted in Table 6.2), independent of whether they are used as a noun to refer directly to a native or as an adjective, used as an attribute. The Italian translations of the “African words” are showed in Table 6.4 below. In the following sections, I will focus on each of them individually.

Table 6.4 “African words”^{*} and their frequency in the TTs

HoD	Trans. S	Trans. B	Trans. M	Trans. G
nigger	5 negro	5 negro	5 negro	5 nero
niggers	5 negri	5 negri	5 negri	5 neri
savage	7 selvaggio	7 selvaggio indigeno	6 selvaggio 1	7 selvaggio
savages	8 selvaggi	8 selvaggi indigeni	7 selvaggi 1	8 selvaggi
negro	3 negro	3 negro giovane di colore	2 negro 1 negretto	2 nero 1
native	1 indigene	1 indigene	1 indigena	1 indigene
natives	5 indigeni	5 indigeni	5 indigeni	5 indigeni
black	22 nera/e/i/o/issime negro/i	14 nera/o/e/i 7 negro/i giovane di colore	16 nera/e/i/o 5 negro/i 1	17 neri/e/a/o 5
blacks	1 negri	1 negri	1 negri	1 neri
Total	57	56	57	57

^{*} *negro/negri*, “nigger”/“niggers”; *selvaggio/selvaggi*, “savage”/“savages”; *indigeno/indigena/indigeni/indigene*, “native”/“natives”; *nero/nera/neri/nere/nerissime*, “black”/“blacks”/“very black”; *negretto*, “little nigger”; *giovane di colore* “young person of colour”

3.1.1 Nigger/niggers

Starting with *nigger/niggers*, all but Translation G of the TTs opt for its direct Italian equivalent, *negro/negri* (“nigger”/“niggers”). In the case of Translation G, a shift is recognisable, for the ST item has been replaced with *nero/neri* (“black”/“blacks”). In line with Kujawska-Lis’s (2008), Trupej’s (2012), and Mouka et al.’s (2015) discussions of the translation of the word *nigger*, I consider translating *nigger/niggers* with *nero/neri* (“black”/“blacks”) to be a shift that mitigates the potential racist

implications of the original term. *Nigger* has been considered “the most socially consequential racial insult” (Kennedy 2002: 32). As a result of its ties with the practice of slavery, *nigger* has been perceived as a “highly offensive racial insult” since 1800 (Hughes 2006: 327). The salience of its social consequences when used as an ethnical slur is witnessed by the fact that *nigger* figures in the reports of the US Supreme Court in episodes concerning racially motivated violence or arson since 1871 (Kennedy 2002: 33). In 1899, when HoD was first published, *nigger* was already changing from “only mildly insulting” to a stigmatising term, a “genuin[e] taboo” (Hughes 2006: 25). *Nero* (“black”), on the other hand, is considered in contemporary Italian a neutral term the choice of which is strongly preferred to *negro* (“nigger”) when used to refer to a black person (*Vocabolario Treccani* 2015: online⁹).

Examining the usage of the two Italian words in the itTenTen corpus, there are noticeable differences. First, *negro* (“nigger”) occurs as a noun 4,655 times (1.51 times per million words, pmw), while *nero* (“black”) occurs as a noun 13,500 times (4.39 pmw): the latter is much more frequently used than the former. Second, *negro* (“nigger”) and *nero* (“black”) display very different collocational relationships. Collocations are calculated setting a span of 5:5 and a minimum frequency of 5, using three different statistics: MI, log-likelihood, and logDice. See Figure 6.6 for a comparison of the collocations of *negro* (“nigger”) and *black* (“nero”).

⁹ Available at <http://www.treccani.it/vocabolario/> (last accessed August 2015).

	Frequency	MI	log likelihood	logDice		Frequency	MI	log likelihood	logDice
P N negro	142	14.035	2,486.949	8.827	P N seppia	497	15.067	9,484.793	9.921
P N Babo	52	16.825	1,119.194	8.427	P N Avola	132	13.406	2,196.853	8.054
P N sporco	265	12.914	4,232.776	8.311	P N affitti	135	11.410	1,868.458	7.478
P N Sporco	35	16.025	711.446	7.841	P N bianco	1,002	10.336	12,437.414	7.432
P N Hoango	16	17.842	370.329	6.802	P N ft	75	13.522	1,260.296	7.361
P N Carmaux	18	13.482	300.802	6.694	P N vestita	125	11.072	1,671.088	7.259
P N zingaro	24	12.338	362.821	6.671	P N Donne	252	10.467	3,159.379	7.255
P N nigger	15	15.807	300.076	6.670	P N avola	55	15.405	1,075.971	7.030
P N merda	137	11.137	1,845.788	6.645	P N vestite	72	11.529	1,008.068	6.952
P N frocio	20	12.664	311.409	6.614	P N corvino	51	15.074	972.081	6.916
P N bianco	14	15.309	270.034	6.554	P N vestito	192	10.023	2,287.845	6.823
P N schiavo	45	11.298	615.421	6.484	P N grigio	155	10.118	1,866.822	6.809
P N ebreo	65	10.976	860.209	6.357	P N illustraz	42	15.358	818.579	6.647
P N terrone	15	12.757	235.492	6.335	P N marrone	59	10.237	719.952	6.294
P N abbronzato	16	12.121	237.013	6.221	P N scuro	87	9.581	982.594	6.177
P N schiaccio	11	13.609	185.761	6.105	P N vestiti	138	9.323	1,509.791	6.174
P N Emmanuel	17	11.519	237.608	6.037	P N rosso	346	9.058	3,663.945	6.134
P N Negro	16	11.625	225.987	6.033	P N emersione	44	9.941	518.812	5.927
P N scappato	23	11.061	306.880	6.000	P N ontano	26	13.199	424.898	5.914
P N bovero	9	18.256	214.957	5.979	P N t.e	24	15.829	485.404	5.854
P N puzzi	9	13.637	152.336	5.847	P N seppie	28	11.398	386.835	5.854
P N Narciso	14	11.381	192.999	5.819	P N dissolvenza	26	12.111	385.079	5.843
P N spiritual	9	13.118	145.814	5.792	P N nero	372	8.736	3,773.920	5.843
P N coglione	21	10.674	268.916	5.709	P N risotto	35	10.098	420.277	5.786
P N spirituals	7	14.407	126.049	5.558	P N lutto	57	9.242	616.873	5.740
P N negri	14	10.740	180.537	5.510	P N spaghetti	39	9.725	448.154	5.735
P N amburghese	7	13.545	117.585	5.508	P N israeliane	36	9.883	421.562	5.732
P N fuggiasco	8	12.197	119.346	5.503	P N tavn	22	13.181	358.943	5.682
P N mulatto	7	13.441	116.564	5.499	P N Attiva	25	10.753	322.932	5.608
P N negra	8	12.006	117.214	5.464	P N Colore	30	9.971	354.940	5.597
P N Atufal	6	17.919	139.656	5.395	P N lucente	24	10.538	302.830	5.517
P N Bernard-Marie	6	16.305	124.392	5.385	P N tinge	22	10.685	282.095	5.445
P N Scontro	9	11.229	122.154	5.369	P N intenso	97	8.526	953.877	5.435
P N grido	49	9.882	573.971	5.364	P N tagliolini	19	11.953	277.215	5.415
P N erculeo	6	14.842	111.735	5.359	P N Belgrado	30	9.281	326.237	5.327
P N Balotelli	17	10.086	203.820	5.215	P N blu	125	8.321	1,193.960	5.319
P N indio	7	11.558	98.203	5.212	P N giallo	95	8.390	916.256	5.317
P N dispregiativo	7	11.545	98.084	5.209	P N lucido	36	8.994	377.165	5.317
P N troia	10	10.618	127.249	5.209	P N argento	83	8.396	801.240	5.287
P N seminaba	5	18.849	125.645	5.135	P N colore	248	8.110	2,298.785	5.219
P N Zumbon	5	17.486	112.750	5.132	P N manodopera	38	8.707	383.043	5.185
P N veluto	5	17.197	110.438	5.131	P N vestirsi	21	9.638	238.729	5.140
P N Masserelli	5	15.984	101.304	5.122	P N sfumature	44	8.506	431.301	5.133
P N semen	5	15.486	97.704	5.115	P N tonalità	33	8.742	334.240	5.122
P N gridandogli	5	14.156	88.270	5.082	P N xilografie	15	12.215	224.346	5.111
P N lurido	6	11.700	85.360	5.071	P N sottopagati	16	11.048	213.242	5.106
P N m.	49	9.509	548.650	5.031	P N Troia	23	9.285	250.227	5.105
P N bastardo	12	10.066	143.518	5.027	P N inchiostro	31	8.762	314.813	5.092
P N xxxxxxxxxx	5	13.125	81.049	5.027	P N lavoretti	17	10.332	209.627	5.077
P N finocchio	8	10.423	99.640	4.946	P N tavole	57	8.183	533.320	5.004

Figure 6.6 Collocates of *negro* (left side) and *nero* (right side)

Excluding proper names (those with the first letter capitalised), the vast majority of the collocates of *negro* (“nigger”) suggest that this word is used derogatively. It occurs frequently with a wide range of insults, from homophobic (*frocio*, “faggot”; *finocchio*, “queer”) to general ones (*sporco*, “dirty”; *coglione*, “asshole”; *troia*, “bitch”; *lurido*, “filthy”; *bastardo*, “bastard”). The left-side table includes also *merda* (“shit”), which refers to the phrase *negro di merda* (literally “nigger of shit”), *schiavo* (“slave”), and *puzzi* (“you smell”). *Negro* (“nigger”) seems also to be related to other ethnical/racial slurs: it occurs with *zingaro* (“gypsy”), *ebreo* (“Jew”), and *terrone* (derogatory term for a southern Italian).

In contrast, *nero* (“black”) does not seem to be used in similar contexts. The collocates in the right-side table refer mainly to the semantic field of colours (*bianco*, “white”; *grigio*, “grey”; *marrone*, “brown”; *rosso*, “red”; *colore*, “colour”; *giallo*,

“yellow”; *blu*, “blue”; *argento*, “silver”), clothes (*vestita/vestite/vestito/vestiti*, “dressed”/“dress”/“dresses”), and art/photography (*dissolvenza*, “fade-out”; *sfumature*, “hues”; *tonalità*, “shades”; *xilografie*, “xylographies”; *inchiostro*, “ink”; *tavole*, “figures”). These collocates reflect the wider variety of uses of *nero* (“black”) as a noun, rather than indicating that it is used as a neutral word to refer to a black person. In this respect, using corpus evidence to support the difference of usage between *negro* (“nigger”) and *nero* (“black”) has some limitations. As acknowledged by Mouka et al. (2015: 64), reference corpora do not always provide evidence for ambiguous terms such as *nero* (“black”); for their analysis, the interpretation of the expression in its specific context of use may be more revealing. Yet, the comparison of *negro* (“nigger”) and *nero* (“black”) in itTenTen has been very insightful, as it has shown that *negro* (“nigger”) is used in very derogatory contexts while *nero* (“black”) is not. This corpus finding can therefore be considered as further support to the argument that the translation shift from *nigger* to *nero* (“black”) mitigates the potential racist implications of the ST word.

3.1.2 *Savage/savages*

Savage and *savages* are translated with their direct Italian equivalents (*selvaggio* and *selvaggi*) in all of the four TTs, apart from one case for each item in Translation B where *savage* is substituted with *indigeno* (“native”) and *savages* with *indigeni* (“natives”). According to Hughes (2006: 147), general terms like *savage* can reflect the prejudicial superiority of the “home” culture over the assumed savagery of outsiders. Similarly, *native* can be seen – especially in the context of colonialism – as a label that generalises and homogenises the colonised other, independently from his/her specific identity and nationality, as opposed to *European*, for example (Hughes 2006: 148). Therefore, both can have derogatory implications, depending on their context of usage, both in English and in Italian.

Using itTenTen to check whether there are marked differences in the usage of *selvaggio* (“savages”) and *indigeno* (“native”) provides little evidence of distinct contexts of use. Therefore, it seems safe to assume that, although replacing *savage/s* with *indigeno/i* (“natives”) can be considered a shift, it is not relevant for this analysis, because my aim is to examine shifts that have the potential to alter the racist

implications of the ST. Moreover, this shift occurs only once in Translation B: as I have argued in Chapter 3, a lexical feature contributes to a style only when it is a repeated choice. A choice in isolation, on the other hand, can be stylistically meaningful *per se*, but it is the patterns of repeated choices that develop the style of a text (Leech & Short 2007: 34).

3.1.3 *Negro*

With *negro*, the translators adopt different strategies. As with *wilderness* and *darkness*, this ST term does not have a direct equivalent in Italian,¹⁰ therefore the translators are faced with a choice here. Translation S replaces it with the Italian *negro* (“nigger”) in all 3 occurrences. Translation B uses *negro* (“nigger”) twice and *giovane di colore* (“young person of colour”) for the remaining one. Translation M adopts *negro* (“nigger”) twice and *negretto* (“little nigger”) once. Finally, consistently with its previous choice, Translation G opts for *nero* (“black”) to replace the ST *negro*.

As for *nigger*, the English term *negro* is stigmatised by its relation with slavery (Hughes 2006: 327). Especially with a lower *N*, as it is used in HoD, *negro* has been “furiously objected” by black people (Kennedy 2002: 114). Today, it is regarded as outdated and offensive (*Oxford English Dictionary* 2015: online¹¹), although until the mid-twentieth century it was recorded in dictionaries as the anthropological designation for “[a] member of a dark-skinned group of peoples originally native to sub-Saharan Africa; a person of black African origin or descent” (*Oxford English Dictionary* 2015: online). Coeval dictionaries in fact focus on its almost “technical” use. *Lloyd’s Encyclopædic Dictionary* (1895) marks it as an anthropology jargon for “[t]he distinctly dark, as opposed to the fair, yellow, and brown varieties of mankind”; *The Chambers English Dictionary* (1905) defines it as “one of the black-skinned woolly-haired race in the Soudan and central parts of Africa [...]”; whereas *A New English Dictionary on Historical Principles* (1888-1928)

¹⁰ To be precise, there are words to translate *negro* in its meaning “member of the Negroid race”, for example *negroide* or *negride*. However, they are much more technical and jargonistic than the English equivalent. In fact, in itTenTen (3 billion words), *negroide* occurs 103 times (0.03 pmw) while *negride* only twice (0.00 pmw), compared to *negro* (“nigger”) and *nero* (“nero”) that occur 4,655 (1.51 pmw) and 13,500 (4.39 pmw) respectively.

¹¹ Available at <http://www.oed.com/> (last accessed August 2015).

defines it as “[a]n individual (esp. a male) belonging to the African race of mankind, which is distinguished by a black skin, black woolly hair, flat nose and thick protruding lips”.

Translating the ST item *negro* with the Italian *negro* (“nigger”) or *negretto* (“little nigger”) can and cannot be seen as a shift that affects the racist implications of the original term. On the one hand, if *negro* (Eng) is considered to be as imbued with racist implications as *nigger* is, then replacing *negro* (Eng) with the Italian term *negro* (“nigger”) would not create much of a shift from the point of view of the racist discourse. Alternatively, if *negro* (Eng) is considered to have been used in its anthropological sense, then replacing it with the slur *negro* (“nigger”) does make a difference. However, the two scenarios are most probably intertwined: even if *negro* (Eng) is used in its technical acceptance, this does not automatically exclude that fact that the word can have racist implications nevertheless. Therefore, the boundaries between the two possibilities are fuzzy.

In contrast, translating *negro* (Eng) with *giovane di colore* (“young person of colour”) and *nero* (“black”) represents a shift that alters the racist implications of the original term. The use of *nero* (“black”) has been already discussed in Section 3.1.1. Replacing *negro* (Eng) with *nero* (“black”) can have an effect similar to substituting *nigger* with *nero* (“black”) (cf. Section 3.1.1). Analogously, translating *negro* (Eng) with *giovane di colore* (“young person of colour”) can be seen as mitigating the racist implications of the ST term. *Giovane di colore* (“young person of colour”) occurs 143 times in itTenTen (0.05 pmw), less frequently than both *negro* (“nigger”) and *nero* (“black”). Looking at its collocates (Figure 6.7), they do not seem to indicate usage in negative contexts, although they are mostly function words and as such they are not very revealing. If the item had occurred more frequently, its description could have been more reliable. A similar limitation is pointed out by Mouka et al. (2015: 64) in their analysis of the translation of racist terms. They too note that low frequency affects the usefulness of their reference corpus in showing the usage of a term.

	Frequency	MJ	log likelihood	logDice
P N un	89	6.077	645.548	-3.282
P N fa	6	5.881	37.371	-3.478
P N era	5	5.116	25.915	-4.243
P N ad	9	5.031	45.862	-4.328
P N una	23	4.678	108.577	-4.681
P N il	6	4.131	23.271	-5.228
P N a	26	4.022	100.647	-5.337
P N che	37	4.018	146.025	-5.341
P N il	32	4.013	124.904	-5.346
P N con	14	3.976	52.195	-5.383
P N in	25	3.898	92.539	-5.461
P N da	11	3.870	39.270	-5.489
P N gli	5	3.847	17.516	-5.512
P N	6	3.800	20.689	-5.559
P N ,	87	3.572	331.744	-5.787
P N al	7	3.488	21.389	-5.871
P N si	9	3.428	26.922	-5.931
P N :	9	3.425	26.887	-5.934
P N di	52	3.412	170.059	-5.946
P N .	46	3.408	148.007	-5.951
P N '	5	3.270	13.843	-6.089
P N del	13	3.255	36.372	-6.104
P N è	12	3.231	33.142	-6.128
P N i	8	3.158	21.177	-6.201
P N e	25	2.932	61.729	-6.427
P N)	6	2.803	13.221	-6.556
P N (5	2.693	10.337	-6.666
P N "	8	2.672	16.474	-6.687
P N non	6	2.534	11.327	-6.825
P N la	10	2.311	16.530	-7.048
P N per	7	2.256	11.044	-7.103

Figure 6.7 Collocates of *persona di colore*

3.1.4 Native/natives

With *native* and *natives*, all of the four TTs translate them with their Italian equivalent *indigeno* and *indigeni*, therefore no shifts occur in the case of this “African word”.

3.1.5 Black/blacks

Despite the availability of quite a straightforward Italian equivalent, *black/blacks* is translated in different ways in the TTs. Translation S uses *nero/i/e/a/issime* (“black”, singular and plural, masculine and feminine, and superlative form) 14 times, omits *black* once, and replaces it with *negro/i* (“nigger”, singular and plural) 8 times. Translation B translates *black/blacks* with *nero/i/e/a* (“black”, singular and plural, masculine and feminine) 16 times, with *negro/i* (“nigger”, singular and plural) 6 times, and once with *giovane di colore* (“young person of colour”). Translation M uses *nero/i/e/a* (“black”, singular and plural, masculine and feminine) 17 times and *negro/negri* (“nigger”, singular and plural) 6 times. Finally, Translation G is the only TT that avoids using *negro* (“nigger”), adhering to the Italian equivalent *nero/i/e/a* (“black”, singular and plural, masculine and feminine) and remaining consistent with the strategy adopted so far.

In line with the discussion in Section 3.1.1 above, the translation of *black/blacks* with *negro/i* (“nigger”, singular and plural) is considered a shift that intensifies the potential racist implications of the ST item. The classification of the shift is less clear in the case of *black* translated with *giovane di colore* (“young person of colour”) in Translation B. Based on the corpus evidence available and their dictionary meaning (cf. Section 3.1.1 and Section 3.1.3), I will not consider this shift as markedly changing the potential racist implications of the ST item.

3.2 Tendencies in the translation of the “African words”

In the previous section, I have compared the translation of each “African word” across the TTs, identifying shifts that have the potential to affect the racist implications that these terms may carry. In this section, I will look at all of the “African words” at once, for each TT, in order to examine whether the individual strategies adopted for the single words create a wider and consistent tendency. Table 6.5 below shows the frequency of the translations of the “African words” for each TT. In this case, however, instead of referring back to the original in order to see how each ST item had been translated, the TTs have been searched for the Italian “African words”, on the basis of the translations listed in Table 6.4. For example, *negro* (“nigger”) has been counted throughout the TTs and all the occurrences have been reported, inclusive of those in which the item is used in contexts where the ST does not use *nigger*.

Table 6.5 Frequency of the translations of the “African words” in each TT*

	Translation S	Translation B	Translation M	Translation G
negro	24	18	19	0
selvaggio	17	17	19	19
indigeno	8	10	7	7
giovane di colore	0	2	0	0
nero	15	16	17	36

* Frequency counts include the frequency of each different form (sing., plur., masc., fem., and superlative)

Table 6.5 shows more clearly the overall tendencies in the translation of the “African words”. Translations S, B, and M have a high frequency of *negro* (“nigger”), higher than the ST (*nigger* and *niggers* occur 10 times in the original). Not only do these TTs consistently use *negro* (“nigger”) in all of the cases in which *nigger* is used in the ST, but they also translate other words with *negro* (“nigger”). In the previous section (cf. Table 6.4), I have shown that it is mostly *negro* (Eng) and *black* that have been translated with *negro* (“nigger”). In fact, *nero* (“black”), in its various forms, occurs less frequently in these three TTs than in the ST (*black/blacks* occurs 23 times in the original). As I have argued with regard to the individual shifts, when reiterated this lexical choice becomes a tendency that can affect the potential racist implications of the ST, intensifying them in translation. In these three TTs, the Africans are more frequently referred to as *negri* (“niggers”) than *neri* (“blacks”), compared to the original.

A diachronic distinction needs to be made, though, as there is a 60 year gap between Translation S and Translations B and M. The use and potential racist implications of *negro* (“nigger”) may have changed in these 60 years. Further research supported by evidence from an Italian historical corpus is needed to prove this hypothesis. However, in the absence of this evidence (an Italian historical corpus is not available), it could be assumed that the usage and potential racist implications of *nigger* in HoD (1902) are closer to the usage and potential racist implications of *negro* (“nigger”) in Translation S (1928) than in Translation B (1989) and Translation M (1990). Therefore, I would suggest that the shifts in Translations B and M can be considered more marked than those in Translation S.

I will discuss some textual extracts and their translations, in order to demonstrate the effects of these shifts. Consider Example (6) below, which shows the translation of *negro* (Eng) with *negro* (“nigger”).

- (6) And every week the messenger, a lone *negro*, letter-bag on shoulder and staff in hand, left our station for the coast.

(6B) E ogni settimana il cursore, un *negro* solitario, sacco della posta in spalla e bastone in mano, lasciava la nostra stazione diretto alla costa.

“And every week the courier, a lone *nigger*, letter-bag on shoulder and staff in hand, left our station for the coast.”

(6M) E ogni settimana il corriere, un *negro* solitario con il sacco della posta sulle spalle e il bastone in mano, partiva dalla stazione diretto alla costa.

“And every week the courier, a lone *nigger* with the post-bag on shoulder and the staff in hand, departed from the station for the coast.”

In the ST, a lone *negro* works in apposition with the subject of the sentence, *the messenger*. If *negro* (Eng) is considered in its anthropological sense (see Section 3.1.3), then Translation B and Translation M can be seen as intensifying the potential racist implications of the original term. However, if *negro* (Eng) is regarded as having similar racist implications as *negro* (“nigger”), then no marked alteration occurs in this case.

In contrast, a marked shift occurs when *black* is translated with *negro* (“nigger”). Consider Example (7) below:

(7) An athletic *black* belonging to some coast tribe and educated by my poor predecessor, was the helmsman.

(7B) Timoniere era un *negro* atletico venuto da qualche tribù della costa, e addestrato dal mio sventurato predecessore.

“The helmsman was an athletic *nigger* come from some coast tribe, and trained by my unfortunate predecessor.”

- (7M) Un *negro* atletico di qualche tribù della costa, istruito dal mio povero predecessore, era il timoniere.
“An athletic *nigger* of some coast tribe, instructed by my poor predecessor, was the helmsman.”

The shift here is more evident than in Example (6). The substitution of *an athletic black* in the original with *un negro atletico* (“an athletic nigger”) in the two TTs intensifies (or even adds altogether) the potential racist implications of the ST item. In the case of Translation B, in Example (7B), the shift is paralleled by another lexical choice: the translation of *educated* with *addestrato* (“trained”). *Addestrare* (“to train”) is commonly used in relation to animals, for instance in the context of animals trained for the circus or trained pets. *Educated*, in the original, does not refer to these contexts, being instead related with intellectual knowledge and schooling.

As far as Translation G is concerned, an opposite tendency compared to Translations S, B, and M is recognisable. As shown in Table 6.5, Translation G replaces every occurrence of *nigger*, *niggers*, and *negro* with *nero* (“black”). *Negro* (“nigger”) does not occur at all in Translation G, while the various forms of *nero* (“black”) occur much more frequently than in the original: 36 versus 23. In line with what I have argued for the individual shifts (see Sections 3.1.1, 3.1.3, and 3.1.5), this consistent lexical choice produces a tendency that mitigates the potential racist implications of the ST. In Translation G, the African natives are always referred to as *neri* (“blacks”) rather than *negri* (“niggers”). See Examples (8) and (9) below:

- (8) A *nigger* was being beaten near by. They said he had caused the fire in some way;
- (8G) Non lontano da lì, stavano bastonando un *nero*. Dicevano che in un modo o nell’altro, era stato lui a provocare l’incendio;
“Not far away from there, they were beating a *black man*. They said that in one way or the other, he had caused the fire;”

(9) And every week the messenger, a lone negro, letter-bag on shoulder and staff in hand, left our station for the coast.

(9G) E ogni settimana, il messaggero della nostra stazione, un *nero* solitario, sacco postale in spalla e bastone in mano, partiva per la costa.

“And every week, the messenger of our station, a lone *black man*, post-bag on shoulder and staff in hand, departed for the coast.”

In Example (8), *nigger* is substituted with *nero* (“black”): this choice mitigates the racist discourse. The syntactic structure of the sentence is modified too. In the ST, the sentence is in passive form and its object, *a nigger*, is placed in the initial position, receiving the emphasis this position possesses. In contrast, Example (8G) shows that the sentence has been converted in an active form: *un nero* (“a black man”) remains the object of the sentence but the initial position is occupied by the implied subject, *loro* (“they”). In this case, then, the emphasis of the violent action of beating someone is put on the perpetrators of such action, namely the whites. Example (9) reports the same extract used in Example (7) to illustrate the different strategy adopted by Translator G compared to the other TTs. Whereas the other translators use the term *negro* (“nigger”) to translate *negro* (Eng), Translation G again employs *black* (“nero”), toning down the potential racist implications of the ST item.

The other shifts described in Section 3.1, related to the translation of *savage/savages* and *native/natives*, do not occur as frequently as those described so far; therefore they cannot be considered as general tendencies. Specifically, the shift from *savage* and *savages* to *indigeno* and *indigeni* (“native” and “natives”) occurs only once in Translation B, as well as the translation of *negro* (Eng) with *giovane di colore* (“young person of colour”) in the same TT. As for the other occurrence of *giovane di colore* (“young person of colour”) in Translation B, it is used as a translation of *black*: I have argued in Section 3.1.5 that this does not represent a marked shift in terms of alterations at the level of potential racist discourse.

Overall, Translation G displays a tendency to mitigate as far as possible the use of potential racist discourse, from a lexical point of view. In this respect, the shifts

in Translation G can be seen as producing a manipulative effect that alters the fictional representation of the African natives in comparison to the original. In contrast, Translation B and Translation M show a tendency to maintain or even intensify the potential racist implications of the ST. These shifts too can be considered to enact a manipulative effect, as they alter – in an opposite way compared to Translation G – the depiction of the Africans. It can be said that the TT reader is more likely to recognise a racist discourse at the lexical level in Translation B and Translation M than in Translation G. Translation S can be seen as the least manipulative one, if the diachronic dimension is taken into account. Even though this translation repeatedly uses *negro* and *negri* (“nigger” and “niggers”), I have hypothesised that the potential racist implications of these terms are equivalent to those of the ST words, therefore this translation choice does not produce a shift as marked as in the other TTs. However, evidence from Italian historical corpora is needed to test this hypothesis and check whether the use of *negro* (“nigger”) at the time is comparable to the use of *nigger* in the original.¹²

The aim of this thesis is not to establish the motivation behind the choices of the translators, but rather to describe the alterations that take place in translation and discuss the potential effect they can have on the representation of the fictional world. However, it is worth pointing out that the manipulation described in this section seems to be of a different nature from that discussed in Chapter 5 (Section 4.4). In Chapter 5, I have argued that manipulation takes place as a result of unconscious alterations of the lexico-semantic patterns of the original, whereas in this case the manipulation seems to be intentional. In Chapter 2, I have outlined the debate on whether HoD is a racist text or not, which has emerged as a consequence of Achebe’s (1990) influential lecture. I have explained that, before Achebe’s (1990) accusation, the theme of race and racism was practically absent from any literary criticism, whereas today the topic is virtually unavoidable. The influence of this debate seems to be reflected in the finding of this analysis, especially in the lexical choices of Translations M, B, and G. The translation of the “African words” in these TTs seems to respond to the debate, as each of the translators alters the fictional representation of the Africans. This outcome shows once more that translation is a social artefact that

¹² As an additional note, it is worth mentioning that a coeval Italian dictionary (*Vocabolario Italiano della Lingua Parlata* 1921) that I have consulted does not list as a meaning of *nero* (“black”) its use as a noun to refer to a black person.

responds to the historical and cultural milieu in which it occurs. As such, it is never a transparent replica of its original, but rather a rewriting that follows dominating canons and norms as much as questioning them. Ultimately, translation enacts a process of production, rather than of reproduction.

3.3 Semantic preferences and prosody in translation

In the previous sections, I have looked at the translations of the “African words” across the TTs and discussed the effects of the translators’ lexical choices on the fictional representation of the Africans in HoD. In this section, I will examine whether the semantic preferences and prosody identified in the ST have been reproduced in the TTs. I will carry out a concordance analysis on the translation of the “African words” (*negro*, *selvaggio*, *indigeno*, *giovane di colore*, and *nero*, in all their different forms) to study whether they enact equivalent local textual functions to those identified in the ST.

The concordance analysis shows that the “African words” in all of the TTs have the same semantic preferences recognised in the ST. Analysing their collocates in a span of 5:5, it is possible to identify the dominant semantic fields recognised in the original: “physicality”, “collectives”, “incomprehensibility”, and “general negative”. This means that in the TTs as well the fictional representation of the Africans is built up around these four aspects. Tables 6.6 to 6.9 provide the words forming part of each of the semantic fields for the four TTs. In addition to the semantic fields, the tables also show that all the TTs do not alter the use of *nero* (“black”) as an adjective premodifying a metaphorical reference to an African, in which the head of the phrase is not a synonym of man or woman. In the TTs too, *nero* (“black”) co-occurs with items such as *ombra* (“shadow”), *cose* (“things”), *forme* (“shapes”), *figure* (“figures”), etc.

Table 6.6 Semantic fields of the “African words” in Translation S

Semantic fields	Words
Physicality	nutrito, piedoni, petto, schiena, corpo, ispalla, ignuda, eretta, testa (2)*, palpebre, gambe, membra, collo, volto, spalla, atletico.
Collectives	ciurma, file, torma, turba, folla (3), fila, campo.
Incomprehensibility	misteriosi, lamentando, gemere, baccano, tumulto, ululati, incomprensibile.
General negative	povero, bastonato (3), idiota, solitario, ferito, polverosi, litigiosa, immusoniti, zoppicanti, lamentando, legnò, servitorello, spietatamente, afflizione, soppressione, indifferenza, morte (2), tumulto, rimpianto, angosciosi, mortale, infelici, odiare, odiarli, cattivi, paura, baruffa, nemici, insolente, immobilità, malattia, svogliatamente.
<i>black</i> [ADJ]	figura nera (2), forme nere (2), nere figure, neri simulacri, così neri, carcame nero.

* Frequency, when ≥ 2 , in brackets

Table 6.7 Semantic fields of the “African words” in Translation B

Semantic fields	Words
Physicality	teste, piedi (2)*, nuda, palpebre, testa (2), ossa, membra, mani, braccia, gambe, collo, petto, pasciuto, atletico.
Collectives	tribù, file, picchetto, banda, masnada, folla, fila, accampamento, torma, massa, turbinio.
Incomprehensibility	gemeva (2), ululato, tumulto, incomprensibile, frenesia, grida.
General negative	pestò, picchiavano, picchiato, gemeva, battuto, solitario, scimunito, impolverati, rissosa, ingrugniti, doloranti, misteriosi, cadavere, doloroso, mortale, infelici, spiacevole, angoscia, morte (2), soppressione (2), indifferenza, rimpianto, odiare, odiarli, rissa, paura, incomprensibile, insolente, indolenti, malattia, inedia, immobilità.
<i>black</i> [ADJ]	figura nera, nere figure, nere ombre, forme nere, così neri.

* Frequency, when ≥ 2 , in brackets

Table 6.8 Semantic fields of the “African words” in Translation M

Semantic fields	Words
Physicality	atletico, teste, musì, nuda, robusto, testa (2)*, palpebre, membra, mani, ossa, braccia, gambe, collo, petto, narici.
Collectives	picchetto, fila, mucchio, file, frotta, tribù, folla, accampamento.
Incomprehensibility	gemeva, lamentava, urlo, grida, frenesia, imperscrutabile, confuse.
General negative	bastonando, bastonato (3), lamentava, gemeva, bastonò, misteriosi, terribili, litigiosa, solitario, idiota, povero, cadavere, dolore, morte (2), soppressione (2), indifferenza, rimpianto, odiare, odiarli, straziato, mortale, infelici, nemici, zuffa, teme va, avvilito, malattia, inedia, insolente, svogliatamente.
<i>black</i> [ADJ]	figura nera (2), ombre nere, forme nere (2), figure nere.

* Frequency, when ≥ 2 , in brackets

Table 6.9 Semantic fields of the “African words” in Translation G

Semantic fields	Words
Physicality	occhi, supernutrito, atletico, nudi, testa (2)*, ossatura, gambe, braccia, membra, mani, piedi (2), corpo, spalla, collo, torace.
Collectives	folla, campo, file, picchetto, fila, banda (2).
Incomprehensibility	ululato, incomprensibile, frenesia, grida, gemeva, lamentava.
General negative	angoscioso, mortale, infelici, odiare, odiarli, morte (2), pena, indifferenza, rimpianto, nemici, paura, impietriti, deficiente, insolente, apatiche, impolverati, misteriosi, litigiosa, immusoliti, doloranti, solitario, malattia, fame, servitore, pietà, bastonando, picchiato, gemeva, bastonato, lamentava, bastonò.
<i>black</i> [ADJ]	ombra nera, sagoma nera, ombre nere, nere ombre, cose nere, forme nere.

* Frequency, when ≥ 2 , in brackets

Tables 6.6 to 6.9 show that, in the TTs as well, the “African words” co-occur with the “physicality”, “collectives”, “incomprehensibility”, and “general negative” semantic fields. In other words, the TTs maintain equivalent semantic preferences as those in the original. This implies that all of the Italian versions have equivalent linguistic features to enact a dehumanising, negative semantic prosody. In translation too, the Africans are described mainly in terms of their bodily appearance, neglecting instead their thoughts and feelings; natives are referred to as indistinct masses, rather than people with their own individuality; they do not communicate in a way that is comprehensible to the colonisers; finally, the Africans are frequently related to negative contexts and situations.

The reproduction in translation of the semantic preferences and prosodies of the “African words” mirrors what has been identified in Chapter 5, where the semantic preferences of the “Africa words” have been similarly maintained in the TTs. In the previous chapter, I have argued that, despite the reproduction of the semantic preferences, translating *wilderness* and *darkness* with multiple terms disrupts the cohesive network existing between the semantic fields and the “Africa words”. The findings in this chapter suggest a similar disruption. Altering the “African words” in translation has consequences on the TTs, even though the original semantic preferences are there. This is the case of Translation G, where contrasting tendencies regarding the representation of the Africans coexist. On the one hand, Translation G mitigates the potential racist implications of the ST at the lexical level (see Sections 3.1 and 3.2 above). On the other, it keeps unaltered the dehumanising, negative semantic prosody produced by the semantic fields. These two aspects, which are in line in the original, are in contrast in Translation G. I will show an example of this phenomenon comparing textual extracts.

The reproduction of the semantic preferences in Translation B and Translation M is not in contrast with the strategy adopted to deal with the “African words”. The increased frequency of *negro* (“nigger”) parallels the reproduction of semantic fields that trigger a dehumanising representation of the Africans. See Example (10) below: the “African word” is in bold; the items belonging to the semantic fields, in italic.

- (10) It was paddled by **black fellows**. You could see from afar the white of their *eyeballs* glistening. They *shouted, sang*; their *bodies* streamed with perspiration; they had *faces* like *grotesque* masks – these *chaps*; but they had *bone, muscle*, a wild vitality, an intense energy of movement, that was as natural and true as the surf long their coast.

(10B) C'erano dei **negri** alla pagaia. A distanza si vedeva risplendere il bianco dei loro *occhi*. *Gridavano, cantavano*, grondavano sudore per tutto il *corpo*; avevano *facce* simili a maschere *grottesche*, quei *tipi*; ma avevano *ossa, muscoli*, una selvaggia vitalità, un'intensa energia di movimento, naturale e vera come la risacca lungo la loro costa.

“There were some **niggers** at the paddle. From afar you could see the white of their *eyes* glistening. They *shouted, sang*, dripped with sweat all over their *body*; they had *faces* like *grotesque* masks, those *types*; but they had *bone, muscle*, a wild vitality, an intense energy of movement, as natural and true as the surf long their coast.”

(10M) Alle pagaie stavano dei **negri**. Di lontano si vedeva brillare il bianco dei loro *occhi*. *Gridavano, cantavano*; i *corpi* grondanti di sudore; quella *gente* aveva *facce* come maschere *grottesche*; ma aveva *ossa, muscoli*, una vitalità selvaggia, un'intensa energia nel movimento, naturale e autentica quanto la risacca lungo la loro costa.

“At the paddles there were some **niggers**. From afar you could see the white of their *eyes* sparkling. They *shouted, sang*; their *body* dripping sweat; those *people* had *faces* like *grotesque* masks; but they had *bone, muscles*, a wild vitality, an intense energy of movement, as natural and authentic as the surf long their cost.”

In Section 3.1, I have explained that Translation B and Translation M use *negro/negri* (“nigger”/“niggers”) when the original employs a different term, an instance of which is shown in Example (10) above. In this passage, *black fellows* is translated, in both cases, as *negri* (“niggers”), in this way intensifying the potential racist implications of the item, compared to the ST. This lexical choice does not clash with the reproduction of the semantic preferences of the “African words” as, generally speaking, they both tend towards a negative representation of the Africans. In Example (10) it is possible

to note the dominating presence of the “physicality” semantic field (*eyeballs, bodies, faces, bone, muscle*), as well as the “incomprehensibility” (*shouted, sang*), the “collectives” (*chaps*), and the “general negative” (*grotesque*) ones. Both Translations B and M translate these terms literally, maintaining the semantic preferences and prosody. Finally, the use of a potentially racist term (*negri*, “niggers”) matches the dehumanising depiction of the paddlers.

In contrast, Translation G mitigates the potential racist implications of the “African words”. However, this tendency does not match the reproduction of the semantic fields, which instead can have dehumanising implications. Translation G applies this toning-down strategy only with the most obvious reflections of the racist discourse, namely the words *nigger*, *niggers*, and *negro*. On the other hand, the dehumanising representation built up by the semantic preferences is not altered, even though this contrasts with the attempt to remove any potential racist inference from the ST. Consider Examples (11) and (12) below.

(11) *Strings of dusty niggers with splay feet arrived and departed;*

(11G) *File di neri impolverati e con i piedi piatti che arrivavano e ripartivano;*

“Strings of blacks, dusty and with splay feet, that arrived and departed;”

(12) *A quarrelsome band of footsore sulky niggers trod on the heels of the donkey;*

(12G) *Una banda litigiosa di neri immusoniti e coi piedi doloranti tallonava l’asino;*

“A quarrelsome band of sulky and footsore blacks heeled the donkey;”

In both examples, the potential racist discourse enacted by the use of *niggers* is mitigated in translation, as the translator replaces it with *neri* (“blacks”). In this respect, Examples (11G) and (12G) are less likely to trigger racist implications than

Examples (11) and (12). However, the semantic preferences of the “African words” is maintained, as well as the resulting dehumanisation: the “physicality” (*feet, foot*), “collectives” (*strings, band*), and “general negative” (*dusty, quarrelsome, footsore, sulky*) semantic fields are reproduced. Thus, when compared to the ST, Translation G shows the presence of contrasting tendencies in relation to the fictional representation of the Africans, a phenomenon that may affect the way the TT reader perceives this particular aspect of the fictional world.

4. Conclusion

This chapter has described and discussed the stylistic tools with which the fictional representation of the Africans is constructed. In particular, this investigation has demonstrated the role that specific textual patterns play in establishing this representation. I have argued that the semantic preferences and prosody identified dehumanise the Africans in the text. This analysis not only has offered linguistic evidence to support what critics have previously hinted at when referring to the way HoD depicts the natives. It has also offered original knowledge towards the appreciation of HoD, providing new insight into how this dehumanising process is actually enacted, which specific lexical features it encompasses, which semantic fields convey the dehumanisation, and how they relate to each other. Therefore, this chapter has contributed new material to the discussion about race and racism in HoD.

From a translational perspective, this chapter has discussed different and, at times, contrasting tendencies in dealing with the construction of the natives in the TTs. First, I have shown how some of the translators’ lexical choices intensify the potential racist implications of the ST. For example, this happens when *black* and *blacks* are translated with *negro* and *negri* (“nigger” and “niggers”). Other choices instead mitigate these implications, for instance when *nigger*, *niggers*, or *negro* are translated with *nero* (“black”). I have argued that these shifts manipulate this specific aspect of the text, with consequences on the reception of the theme “race and racism” in the target context. Second, I have discussed how the reproduction of equivalent semantic preferences and prosody can be in contrast with the way the “African words” have been translated. This is the case of Translation G, where contrasting tendencies

in the depiction of the Africans coexist. Overall, the analysis of the translations has shown how alterations at the micro-linguistic level can have effects at the macro-textual, interpretational level.

This chapter is related to the previous one in the discussion of translation manipulation. The analyses in Chapter 5 and Chapter 6 have shown different types of translational shifts. In the previous chapter, I have argued that the shifts identified are the result of the lack of a translation equivalent and are therefore obligatory. Most of the shifts identified in this chapter are non-obligatory, as an equivalent is available. As I have demonstrated, both shifts can create manipulative effects: this provides further evidence that, independently from their intentionality or motivation, both types of shifts have the same potential to manipulate the TT.

Finally, before moving on to the next chapter, it is worth pointing out another parallelism existing between the two analyses. In Chapter 5, I have shown that Translation G is the TT that displays a different strategy in dealing with the fictional representation of the African jungle compared to the other TTs. In contrast, Translations S, B, and M show a more similar behaviour in the translation of *wilderness* and *darkness*. The analysis reported in this chapter has shown a comparable picture. Translations S, B, and M show equivalent tendencies in dealing with the “African words”. In particular, Translation B and Translation M share this tendency even more closely, given the similarity of the socio-cultural contexts in which the translation practice took place. In contrast, Translation G again displays an independent behaviour, differing markedly from the strategies adopted by the other TTs. This outcome will be paralleled and further explored in the next chapter.

CHAPTER SEVEN

Applying PCA to the study of literary translation

1. Introduction

In this chapter, I will use principal component analysis (PCA) to compare the TTs and the ST. The aim is that of complementing the comparison of the translations with a perspective different from that provided by the previous chapters. Whereas in Chapter 5 and Chapter 6 I have compared the TTs based on specific textual features, in this chapter I will compare the translations as whole texts, on the basis of the frequency patterns of items that have not been taken into account previously. PCA will use the frequency patterns of such a large portion of the total number of tokens in the texts that, statistically speaking, the analysis can be said to approximate comparison at the level of whole texts. The outcome of this analysis will contribute to a more nuanced understanding of the findings of the previous chapters, as the alterations of the specific textual features identified in Chapters 5 and 6 will be related to the overall picture of the whole texts' relations that PCA offers. At the same time, I will examine whether the differences already identified at the local level are equally reflected at the level of entire texts. This twofold perspective will provide a more comprehensive picture of how the ST and the TTs relate to each other, a picture that combines different levels of analysis. Finally, the convergence of these levels will complete the methodological cycle described in Section 5 of Chapter 3.

The input of PCA is parallel lists: in Section 2, I will introduce parallel wordlists and parallel word-sequence lists, illustrating how they have been obtained and how I will use them. In Section 3, I will start the analysis by comparing the TTs to each other. This comparison will show whether and to what extent the translations differ when compared among themselves as whole texts, without referring back to the ST. The comparison will be established in two different ways. First, in Section 3.1, I will use the frequency patterns of the most frequent words in the TTs to calculate the overall degree of similarity between them. Second, in Section 3.2, I will repeat the analysis using the most frequent sequences instead. In Section 3.3, I will discuss the results of the two comparisons and relate the findings to each other.

In Section 4, I will compare the TTs in relation to the ST. Given that the language difference between the original and the translations prevents a direct comparison, I will propose an alternative method to compare the TTs to the ST: a procedure based on the division of the texts into sequential segments. I will first apply this procedure to HoD, in Section 4.1. Then, in Section 4.2, I will compare the TTs based on the application of the procedure on the original, examining the differences and the similarities in the results. This analysis will provide different findings from the analysis in Section 3. I will therefore discuss how these findings differ in Section 4.3. In Section 5, I will explain why the findings differ, introducing the notion of ‘stylistic overtones’ (McKenna et al. 1999) to interpret the outcome of my analysis. Finally, I will discuss the contributions of this chapter, in relation to the previous chapters, in Section 6.

2. Parallel word- and word-sequence lists

PCA is a statistical method that converts the observations of a large amount of variables into a smaller number of values, i.e. principal components. The principal components are defined so as to account for as much variation in the data as possible. It is therefore a procedure that facilitates the handling of a large quantity of data by compressing it into a more manageable format. In corpus/computational linguistics/stylistics, PCA is usually used to obtain a measure of difference across texts or corpora based on the performance of a given number of variables (see Chapter 3, Section 4.4).

Generally speaking, these variables are the most frequent words (MFWs) or most frequent sequences (MFSs) in the texts under analysis. PCA uses the frequency patterns of these MFWs/MFSs to provide a statistical approximation of the overall degree of similarity of a number of text-specimens. The more similar the patterns are across texts, the more likely the relation of similarity between these texts; the more different the patterns are, the higher the ‘degree of unlikeliness’ (Burrows 2002b). The frequency patterns necessary for the calculation are provided in the form of parallel lists: wordlists when using MFWs as variables, and word-sequence lists when using MFSs. A parallel list includes the frequency list of all the items (words or word

sequences) in a corpus, plus similar lists for each text in the said corpus (cf. Hoover 2012: online). The corpus word- or word-sequence list is sorted by decreasing frequency, while the text lists are not; rather, they follow the order set by the corpus list. See for example an extract of a parallel wordlist in Table 7.1 below.

The first column lists the words in the corpus including the four TTs in descending frequency order. The other columns are not sorted by frequency: the frequency of *che* in Translation S, for instance, is higher than the frequency of *e*, yet *che* comes after *e*. The order that the translation columns follow is dictated by the corpus wordlist (the first column). Next to each of the entries in the “Word” column is the frequency of that relative word in each translation. When an item in the “Word” column does not appear at all in a translation, the frequency for that word in the relative TT column would be 0. In Table 7.1, all the frequency entries are normalised so that they can be compared: the entry represents the word frequency expressed as a percentage of the total number of running words in that translation.

Table 7.1 Sample of TTs parallel wordlist

Word	Translation B	Translation G	Translation M	Translation S
di	3.7231503	3.7494154	3.7497299	4.311933
e	2.5536993	2.45284	2.5529501	2.3330383
che	2.2036595	2.5827575	2.3422303	2.421574
un	2.3945904	2.2371771	2.2881997	2.1942523
a	1.8615751	1.6343606	1.7532959	1.9908593
la	1.7077699	1.9123837	1.8181328	1.4572515
il	1.7157253	1.7408928	1.769505	1.3711087
non	1.4823654	1.6031804	1.5047547	1.636716
in	1.4478918	1.3251572	1.3642749	1.3735014
una	1.286131	1.22642	1.212989	1.3447871
era	1.2277911	1.3199605	1.253512	1.1078941
per	1.0501193	1.0289456	0.9725524	1.0432868
si	0.9573057	1.0653224	1.0076723	0.71067935
con	0.9626094	0.8314712	0.8617895	0.74417937
mi	0.8035004	0.8470613	0.86449105	0.8135723
l'	0.7186423	0.83926624	0.8725956	0.6412864
le	0.65234685	0.69635713	0.62945753	0.5958221
come	0.73190135	0.62620175	0.6456667	0.55035776
[...]	[...]	[...]	[...]	[...]

Parallel word-sequence lists work similarly to parallel wordlists: they have the same ordering and format of a parallel wordlist, conveying the same frequency information but on word sequences instead of individual words.

In Section 4, I will divide the texts into sections and use them to compare the ST and the TTs. Again, the functioning of a parallel list does not change. In this specific case, the main wordlist is that of the text under examination, while the parallel lists are those of its constituent sections. For instance, referring to the sample in Table 7.1, the first column would show the word occurring in the text sorted in the usual descending frequency order; the following columns would show the normalised frequency of the corresponding words as they occur in each section of said text.

It is important to specify that Italian words with apostrophes are considered as autonomous tokens in this study, instead of parts of the items they occur with. For example, *l'albero* (“the tree”) counts as two separate tokens: *l'* and *albero*. In fact, *l'* appears on its own in the parallel wordlist in Table 7.1. Apostrophes are commonly used in Italian. One of their main uses is to indicate the dropping of a vowel at the end of a given word, when such a word precedes another word starting with a vowel. For example, “*lo + albero*” (“the + tree”) becomes *l'albero* (“the tree”). Elided words are generally very common function words (such as *l'*, definite article, and *d'*, preposition), whose frequency is much higher than that of the word they occur with (often, content words). Considering an elided word and the word it occurs with as a single token would mean substantially affecting the frequency of the elided function word. The total frequency of the elided word would then be split into the several lower occurrences of the other content words it occurs with. For example, *l'albero* (“the tree”), *l'orologio* (“the watch”), and *l'occhio* (“the eye”) would be counted as three separate types in the case elided words were considered part of the word they occur with. In contrast, when considering them individually, we would have four types: *l'* occurring three times and the other three nouns each occurring once. Tognini-Bonelli (2001: 138) adopts the same distinction in her study of *in case of* and its Italian translation *nel caso di*. She considers the elided form of the preposition *di* merged with the definite article *lo (dell')* as independent from the word it follows.

Laviosa (personal communication)¹³ agrees with Tognini-Bonelli (2001) and suggests that words with apostrophes should be counted separately.

Parallel wordlists are generated with *Intelligent Archive*¹⁴ (cf. Chapter 3, Section 4.5). However, the application does not process word sequences, so parallel word-sequence lists have been created using Hoover's (2012: online) *The Parallel Wordlist Spreadsheet*, a Microsoft Excel spreadsheet with macros that is able to process word sequences as well as single words.¹⁵ Parallel word- and word-sequence lists are then exported to *Minitab 16* to be processed.

3. Comparing the TTs to each other

In this section, I will examine the overall degree of similarity between the translations, independent of how similar or dissimilar they are from the ST. The TTs will be compared to each other on the basis of the frequency patterns of MFWs first (Section 3.1) and then MFSs (Section 3.2). In the case of MFSs, I will use two types of sequences: two-word sequences, in Section 3.2.1, and three-word sequences, in Section 3.2.2. Finally, in Section 3.3, I will discuss the findings obtained.

3.1 Comparing the TTs to each other using MFWs

In this section, I will analyse the overall degree of similarity between the translations based on the frequency patterns of their MFWs. But how many MFWs to take into account? There is no general agreement on this number and different studies use different numbers. For example, McKenna et al. (1999) use 99 MFWs first and then 61 and 65 when inflected verbs and inflected personal pronouns are removed from their list. Burrows (2002b) uses 20 MFWs in his analysis of the translations of Juvenal's "Tenth Satire", while he employs pools of 150, 120, 100, 80, 60, and 40

¹³ Private email, 24/02/12.

¹⁴ Retrieved from <http://www.newcastle.edu.au/research-and-innovation/centre/education-arts/cllc/research> (last accessed April 2015).

¹⁵ Retrieved from <https://files.nyu.edu/dh3/public/TheParallelWordlistSpreadsheet.html> (last accessed April 2015).

MFWs when testing his Delta measure (Burrows 2002a). Grabowski (2013) starts with 1,000 MFWs and then decreases the number to 950 MFWs after deleting the top 50 MFWs so as to exclude the most frequent function words. Rybicki et al. (2014) too use 1,000 MFWs in their analysis of the works written collaboratively by Conrad and Ford. The reasons why the studies cited here use these specific pools of MFWs are not explicitly stated.

Rybicki (2006) is more explicit about this methodological decision. His choice of using 200 and 250 MFWs is the result of a series of preliminary tests with different numbers of MFWs (30, 100, 200, and 250). After comparing the tests, Rybicki (2006: 93) explains that the results were more stable with larger numbers, hence the decision to use 250 MFWs for the Polish original and 200 MFWs for the English translations. A similar method is applied in Rybicki & Heydel (2013: 711): their analysis is performed with different ranges of MFWs (from 100 to 1,000) and then the individual results are combined in order to produce the most consistent outcome.

Hoover (2002: 159) explains that increasing the number of MFWs usually increases the accuracy of the analysis, but only until a certain point, after which increasing the number of MFWs in fact reduces accuracy. He suggests that a reason behind this phenomenon may be that very large pools of MFWs lead to the increase of the number of words that do not occur in some of the texts. The larger the number of words taken into account, the larger the proportion of words that appear only in some texts and not in the others. This, Hoover (2002: 160) suggests, can affect the accuracy of the analysis. To prove his hypothesis, Hoover (2002) carries out a cluster analysis with different pools of MFWs, ranging from 20 to 500. He shows that accuracy rises with the increase of MFWs, up to 88% when 300 MFWs are used. After this peak, accuracy decreases as more MFWs are added. In the specific case of PCA, Hoover (2015: online) proposes a similar approach, i.e. starting with the largest number of MFWs possible and then decreasing the number gradually: “what you are looking for, and what you will usually find, is a series of analyses at some points that are very similar. That should be the most reliable results [...]” (Hoover 2015: online).

For the study in this chapter, I have followed Rybicki’s (2006) and Hoover’s (2002, 2015) approach, consisting of carrying out preliminary analyses with different numbers of MFWs and then selecting the number of MFWs in which results stabilise. Specifically, I have started from 995 (the largest number of variables possible in

Minitab 16) and then decreased the number to 850, 750, 650, 550, 450, 350, 250, 150, and 50. The results were different with the largest numbers of MFWs; they started to stabilise from the threshold of 750 MFWs on, and became stable in the range of 350-250 MFWs. With smaller pools of MFWs (150 and 50), the results were again different. I have therefore decided to use 250 MFWs for the present study.

Figure 7.1 below shows the score plot representing the overall degree of similarity between the TTs, based on the frequency patterns of the 250 MFWs. As mentioned before, PCA is a procedure of dimension reduction and a score plot is the visual representation of this reduction: the score plot allows me to concentrate high-dimensional information (the frequency patterns of 250 words across 4 different texts) and visualise it in a low-dimensional space, a 2D graph. The two axes are the two principal components that account for as much variation across the data as possible. Each red dot stands for one of the TTs; the variation between them is represented by their distance: the closer the translations are to each other, the more similar they are; the further away they are, the more variation there is among them. The first component, the horizontal axis, always represents more variation than the second component, the vertical axis. Therefore, distance between the TTs on the horizontal axis represents stronger variation than distance on the vertical axis.

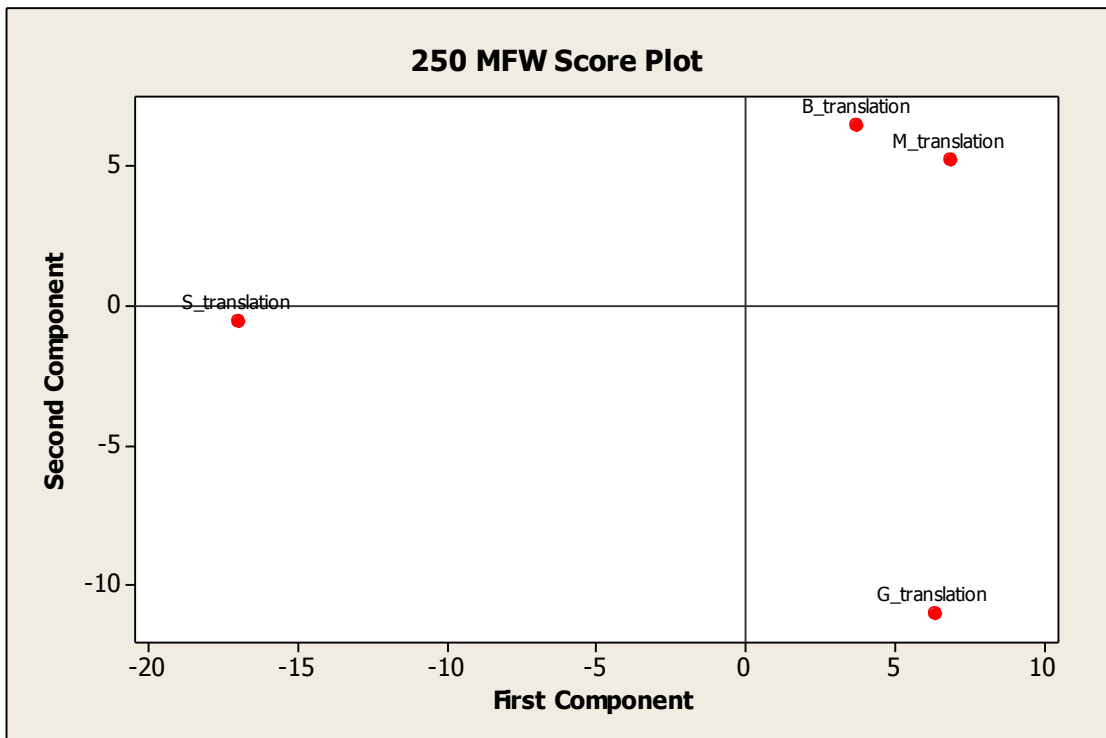


Figure 7.1 Score plot based on the 250 MFWs of the TTs

As the strongest divergence is that represented on the horizontal axis, the sharpest difference is that between Translation S and Translations B, M, and G. However, the latter do not cluster all together: Translation B and M form a close pair, separated markedly from Translation G. The distance between Translation S and Translations B, M, and G can be interpreted as a reflection of the temporal distance between them. In fact, there is a 62-year gap between the two sets. In contrast, there is no temporal gap between the contemporary TTs, therefore the same interpretation cannot be used to explain their relations. The closeness between Translation B and Translation M, and their distance from Translation G is more probably due to the translations' individual styles.

The score plot does not provide any data to verify the reasons behind this particular clustering, but simply shows the existing divergences. Yet, it is important to note that this configuration matches the findings of the previous chapters. Both the study of the fictional representation of the African jungle (Chapter 5) and that of the fictional representation of the African natives (Chapter 6) have shown an analogous picture: Translations B and M behave similarly in respect to the translational issues discussed in Chapters 5 and 6, as opposed to Translation G, which displays instead a

hand side of Figure 7.2 represent the strongest word-discriminators of Translation S, namely those items whose frequency patterns discriminate most markedly Translation S from the other TTs. The words that lie at the right-hand side of Figure 7.2 represent instead the strongest word-discriminators of Translations B and M, and Translation G. The large number of words and the fact that they overlap each other make it difficult to read them clearly in the figure. However, *Minitab 16* provides a set of options, from simple zoom to formatting word labels, which allows me to have a clearer picture of what happens inside the clusters of words. Here, I will describe the most relevant patterns identified; the interested reader can find larger and more readable version of the scatterplots in Appendix 2.

Figure 7.2 confirms the interpretation according to which the divergence between Translation S and Translations B, M, and G is due to their temporal distance. Translation S differs from the contemporary TTs for the use of archaic or dated forms that are not commonly adopted in (literary) Italian today. For example, among the discriminators of Translation S it is possible to identify *m'*, *s'*, and *d'*. *M'* and *s'* are the elided forms of the object personal pronouns *mi* and *si* (first person singular and reflexive third person singular, respectively), as in *voi m'intendete* and *la fucileria s'arrestò* (from Translation S), as opposed to *voi mi intendete* and *la fucileria si arrestò* (“you understand me” and “the fusillade stopped”). Likewise, *d'* is (in most of cases) the elided form of the preposition *di* (“of”/“at”/“in”), as in *una tavola d'abete* (Translation S), as opposed to *una tavola di abete* (“a spruce desk”). These elided forms sound dated today and their use is currently decreasing in written language. The fact that they appear as what discriminates Translation S from the contemporary TTs is a confirmation of this tendency.

Other examples of dated forms among the discriminators of Translation S are *ed*, *ad*, and *egli*. *Ed* (“and”) and *ad* (“at”/“in”/“to”) are the euphonic forms of *e* and *a*, as for example in *ed alberi immensi* and *in faccia ad un pericolo* (both from Translation S), as opposed to *e alberi immensi* (“and immense trees”) and *in faccia a un pericolo* (“in the face of danger”). The addition of a euphonic *-d* is nowadays avoided in literary language, while it is kept in the case of fixed expressions (e.g. *ad un tratto*, “all of the sudden”) or when *e* (“and”) or *a* (“at”/“in”/“to”) are followed by a word starting with the same vowel (*e-* or *a-*). For instance, when *egli* (“he”), another discriminator of Translation S, co-occurs with the conjunction *e* (“and”), it becomes

ed egli (“and he”), namely in the use of the euphonic *-d*. *Egli* itself is a dated form, the use of which has dramatically dropped both in written and spoken language, in favour of the alternative *lui* (“he”). This decrease in use is confirmed by the fact that the other TTs use *lui* (“he”) instead of *egli* (“he”).

Egli is not the only pronoun among the discriminators of Translation S. Also *voi* (“you”, subject personal pronoun second person plural) and *vi* (direct object pronoun second person plural or reflexive pronoun second person plural) contribute to distinguish Translation S from the other TTs. These discriminators too mark the temporal distance between the texts, for in Translation S *voi* and *vi* are used as courtesy forms to address people in formal situations. This practice was particularly widespread at the time Translation S was written, during the fascist regime, when the use of *voi*, as opposed to *lei*, was eagerly fostered (Ajello 2008).¹⁶ Today, also as a reaction against the regime imposition, *lei* (“she”) is very much preferred to *voi* as courtesy form, and much more widely used. In fact, all the TTs apart from Translation S use *lei* as a polite form to address people in formal contexts.

The differences between Translation S and the other TTs do not seem to be limited to the influence of the temporal gap only, since there are other discriminators that cannot be related to diachronic changes. An example is the group of demonstrative deictics clustering around Translation S: *quell*, *quel*, *quella*, *quella*, and *quei* (“that” and “those”, masculine and feminine). Their presence signals a higher frequency of use of this grammatical class in Translation S compared to the other translations. The function of deictics is that of identifying a specific element that is being referred to; demonstrative deictics, in particular, refer to the degree of proximity to the speaker/writer (Halliday & Matthiessen 2004: 314).

Demonstratives are not the only type of deictics that cluster around Translation S. Person deictics can be found too: apart from the already mentioned *egli*, *m'*, *s'*, *voi*, and *vi*, there are also *io* (“I”), *me* (object personal pronoun first person singular), *suoi* (“his”), and *sue* (“her”). There are some more spatial deictics as well, such as *là* (“there”) and *laggiù* (“down there”). The use of deictic devices helps to establish the spatial-temporal coordinates of what is being narrated (Wales 2011: 107). This assists the process of involvement into the narration and allows the reader

¹⁶ Retrieved from <http://ricerca.repubblica.it/repubblica/archivio/repubblica/2008/01/27/la-guerra-dei-pronomi-nell-italia-in.html> (last accessed August 2015).

to share the narrator view-point more intimately. This process seems to be more marked in Translation S than in the other TTs, based on the presence of deictics among the discriminators that distinguish the early translation from the contemporary ones.

Moving on to the contemporary translations, the scatterplot confirms what has been noted in Figure 7.1, namely, the higher degree of similarity between Translations B, M, and G. This is indicated by the fact that the words on the right-hand side of Figure 7.2 cluster more densely between the relative positions of the contemporary TTs. Only a smaller number of stronger discriminators lie at the two ends of the cluster, on the vertical axis, closer to the relative positions of Translation G, and Translations M and B. This means that the items specific to each contemporary translation are fewer than those used in a similar way by all three of them, especially in the case of Translations B and M. In Figure 7.1, they cluster so closely together that it is practically impossible to distinguish the discriminators of one translation from those of the other in Figure 7.2. Their degree of similarity is certainly higher than their degree of divergence.

Among the discriminators that refer to the contemporary TTs, it is possible to identify some groups of words that may indicate translators' idiosyncratic preferences. These groups gather together words belonging to the same grammatical category or word class, suggesting a preference for a specific syntactic structure or construction, rather than the preference for an individual word. For example, several forms of the preposition *al* ("to"/"at") are related to Translations B and M: *alle*, *alla*, *al*. On the other hand, Translation G appears to prefer the preposition *dei* ("of the") and its various forms, as confirmed by the relation between the position of *dei*, *delle*, *del*, and *degli* on the scatterplot and the position of Translation G on the score plot.

In Figure 7.2 it is possible to identify other examples of groups of words belonging to the same grammatical category/class clustering together. For instance, the demonstrative deictics described above in relation to Translation S: *quell*, *quella*, *quel*, *quelle* and *quei* ("that" and "those", masculine and feminine) are all different forms of the same lemma. Another instance is the preposition *su* ("on"/"above"/"up"/etc.) and many of its forms, all appearing in the large cluster between Translation B and M, and Translation G: *sulla*, *sulle*, *su* and *sul*. In Chapter 3, I have argued that function words can be seen as the markers of the relationships

that words establish with each other in a text. As such, function words are also the markers of everything these relationships entail (McKenna et al. 1999: 152). The use of specific function words – for example the prepositions just described – leads to the use of specific phrase and sentence constructions, e.g. prepositional phrases. The fact that different forms of the same function word occur close together on the scatterplot indicates that they are used more frequently by one translator compared to the others. In turn, this distinctive use of words belonging to the same category or class can be seen as a preference for a given construction rather than another. In this respect, the scatterplot indicates that it is mainly these stylistic idiosyncrasies that distinguish Translations B and M from Translation G.

3.2 Comparing the TTs to each other using MFSs

In this section, I will compare the TTs to each other by investigating their overall degree of similarity based on the frequency patterns of MFSs. Within the aims of the present chapter and in agreement with Hoover (2002: 16), I will adopt a working definition of ‘word sequence’ as a sequence of two or three contiguous words. Shifting the focus from single words to word sequences allows me to widen the scrutiny to more complex lexico-grammatical structures, improving the understanding of the relations between the TTs.

The MFSs used in this section are two-word and three-word sequences. Barlow (2013) also takes into account two-word and three-word sequences, in his multifunctional analysis of the spoken output of six White House press secretaries. Conversely, Hoover (2002) uses only two-word sequences and ignores longer sequences. Hoover (2002: 162) claims that three-word (or longer) sequences do not occur frequently enough to be worth investigating. To bear this out, he carries out a cluster analysis with 40 three-word MFSs, proving that only in 6 out of 17 cases the author is attributed correctly. However, my analysis does not aim at attributing correctly texts to their authors. Therefore, in line with Barlow (2013), I will use two-word (Section 3.2.1) and three-word sequences (Section 3.2.2) and show that in both cases the analyses provide relevant results, as they are consistent with what has been obtained using MFWs.

As for the investigation in the previous section, I have carried out preliminary analyses with different pools of MFSs in order to identify the number of MFSs in which results were more stable (see Section 3.1 above). Working with two-word MFSs, the results were very similar from 990 to 350. 250 MFSs produced slightly different results, while 150 and 50 MFSs gave very different results to those obtained before. Therefore, 350 is the number of two-word MFSs used, as again this pool produces results that have been stabilising for the six previous analyses. With respect to three-word sequences, the results were basically the same from 990 to 150, changing only when 50 MFSs were used. For the sake of consistency, 350 is the number of three-word MFSs taken in consideration.

3.2.1 *Comparing the frequency patterns of two-word MFSs*

To start with, the TTs are compared using the 350 two-word MFSs as variables. The results (see Figure 7.3 for both the score plot and scatterplot) show an unequivocal similarity with what has been obtained using MFWs. In Figure 7.3 too, the distance between Translation S and the contemporary translations represents the strongest variation among the texts, while Translations B, M, and G cluster close together, signalling their similarity. However, it is still possible to recognise a closer relation between Translation B and Translation M, while Translation G is the most different among the contemporary TTs. Overall, the relations among the TTs are unaltered.

The clustering of the discriminators on the scatterplot confirms what the score plot shows. Translation S has its own cluster of sequence-discriminators on the left-hand side of the plot, which is separated from the other large cluster on the opposite end of the graph, gathering more homogeneously the discriminators of Translations B, M, and G (see Appendix 2 for a larger scatterplot).

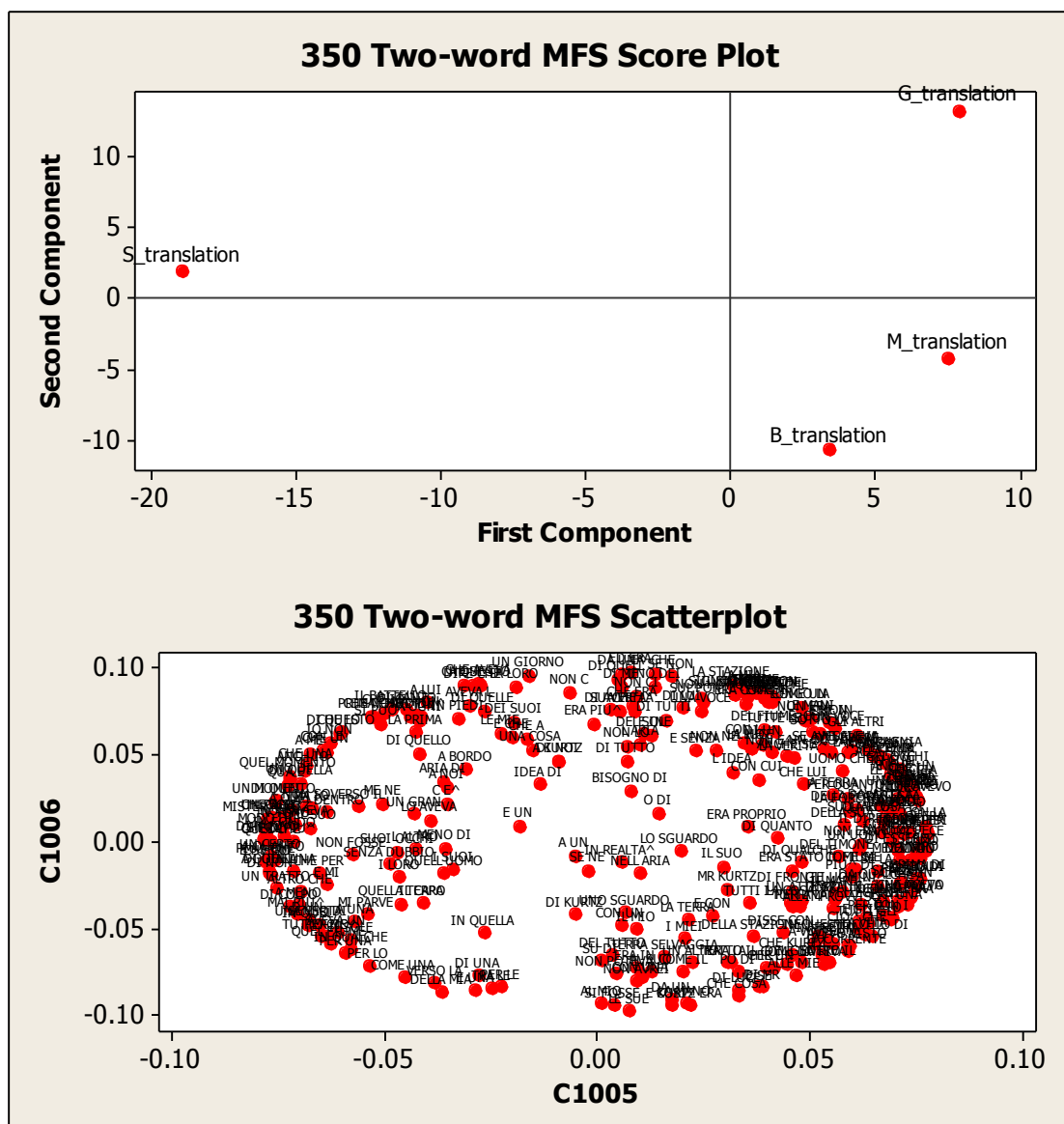


Figure 7.3 Score plot and scatterplot based on the 350 two-word MFSs of the TTs

A look at the MFSs in the scatterplot suggests that, this time, the temporal gap is not the strongest discriminating factor. Examining the MFSs that characterise Translation S, dated and archaic forms do not abound. Only a very few of the MFSs refer back to the old-fashioned MFWs identified in the previous section, for example the MFS *s'era* (reflexive third person pronoun + “was”) which refers to the MFW *s'*. The vast majority of the MFSs do not seem to be related to archaic or disused items. Looking only at Figure 7.3, it would be very difficult to guess that Translation S is 68-year older than the other TTs without being already aware of the temporal gap.

In contrast, there is a clear abundance of MFSs that refer back to the MFWs that were not linked to the temporal gap, especially deictics. There are many MFSs

with person deictics (*io non*, “I do/did not”; *a me*, “to me”; *anch’io*, “me too”; *ma io*, “but I”; *che egli*, “that he”; *s’era*, reflexive third person pronoun + “was”; *del suo*, “of his”; *che gli*, “that” + indirect object third person pronoun; *e mi*, “and” + indirect object first person pronoun; *mi parve*, “seemed to me”; *per me*, “for me”), determiner deictics (*quel momento*, “that moment”; *di quella*, “of that”; *di quei*, “of those”; *a quel*, “at that”; *a quella*, “to that”; *di quel*, “of that”), and spatial deictics, such as *qua e* (“here and”) and *là dentro* (“there inside”).

The right-hand side of the scatterplot is more difficult to interpret because, similarly to what has been seen in Figure 7.2, Translations B, M, and G are more similar to each other and their discriminators tend to merge together. Most of these MFSs cluster together between Translations B and M, and Translation G; only a smaller portion of them cluster closer to the translations’ relative position. The higher homogeneity between Translation B and Translation M makes it even more difficult to distinguish the discriminators of one TT from those of the other. Although it is possible to recognise some MFSs related to the MFWs discussed before, it is virtually impossible to separate the discriminators of the two TTs because of their high homogeneity and the density of items in this area.

3.2.2 Comparing the frequency patterns of three-word MFSs

Switching from two-word to three-word MFSs (Figure 7.4) does not change the overall picture. Translation S is in the left-hand side of the score plot, far away from the other three TTs. In the right-hand side, Translations M and B cluster close together, separated on the vertical axis from Translation G. The scatterplot shows the same disposition of clusters as in Figure 7.3: on the left-hand side, there are the discriminators of Translation S, while, on the right-hand side, there are those of Translations B, M, and G (see Appendix 2 for a larger scatterplot).

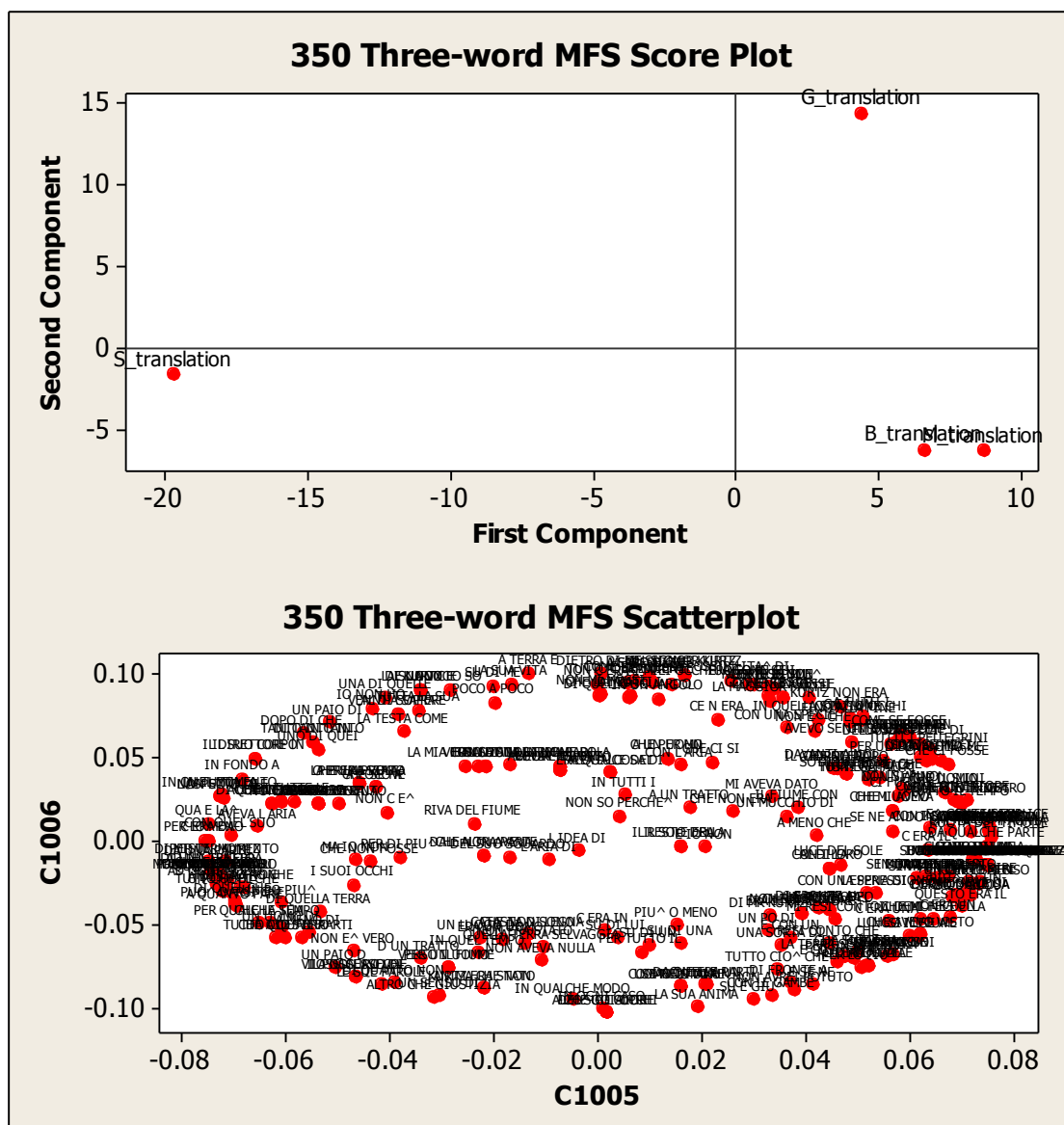


Figure 7.4 Score plot and scatterplot based on the 350 three-word MFSs of the TTs

Looking at the individual sequences confirms what has been noted in the two-word MFS analysis. The linguistic variation due to the temporal gap is hardly recognisable in the discriminators of Translation S. At the same time, there are many three-word sequences (*con quel suo*, “with that [...] of his”; *in quel momento*, “in that moment”; *quanto a me*, “for what concerns me”; *che egli non*, “that he do/did not”; *da queste parti*, “here around”; etc.) that refer back to the features identified in the previous analysis as characteristic of Translation S, mainly deictics.

Focusing on the right-hand side of the scatterplot, most of the discriminators are again located in a dense cluster between Translations M and B, and Translation G,

whilst the MFSs closer to the relative positions of the TTs are fewer. This makes it difficult to assign the discriminators to any specific translation. The smaller clusters, those closer to the relative position of each translation, do not offer much information on the differences between the contemporary TTs, as only a few sequences that refer to items highlighted in the MFW and two-word MFS analyses are recognisable.

Overall, the use of MFSs has shown a high degree of consistency with the previous findings. Such a consistency across the analyses using MFWs, two-word, and three-word MFSs can be seen as an indication that the divergence between the TTs results from differences at the level of text construction. Function words – most of the MFWs taken into account – act as building agents for the structuring of sentences; they establish those relationships among content words which help the text to convey meaning. Compared to single MFWs, two-word and three-word MFSs are one step ahead in the process of constructing the text, the latter just adding one level more to the former. The consistency in the findings mirrors the consistency in the choices of the translators, suggesting that MFW, two-word MFS and three-word MFS analyses refer to the same process of sentence construction at distinct levels of development. Ultimately, it is this very process of sentence construction, with its ideolectal differences and the choice of using given lexico-grammatical tools instead of others, which distinguishes the TTs from each other.

3.3 Relation between the findings

In the previous sections, I have analysed the relations between the TTs, based on the frequency patterns of their 250 MFWs and 350 two-word and three-word MFSs. Before moving to the next section, where I will compare the TTs in relation to the ST, I will discuss the findings obtained so far.

Using MFWs as variables, the analysis has shown that the most marked variation is exhibited between Translation S and the contemporary TTs, and that the strongest discriminating factor is the temporal gap between the TTs. The examination of the individual discriminators has confirmed that most of the MFWs that distinguish Translation S are in fact archaic and dated forms dismissed in contemporary (literary) Italian. However, other types of discriminators that do not refer to the temporal gap

have been identified too, such as deictics. On the other hand, the contemporary TTs display a higher degree of homogeneity, especially in the case of Translation B and Translation M. The investigation of the individual discriminators has indicated that each of the three contemporary TTs has idiosyncratic preferences for specific function words, clustering together with related forms (singular and plural, masculine and feminine) of the same grammatical category/class. These preferences seem to be what distinguishes the contemporary TTs.

I have suggested that the preferences for a given grammatical category/class can indicate ideolectal differences at the level of sentence structuring. Further evidence to this has been provided by analyses based on MFSs. Using two-word and three-word sequences has allowed me to widen the perspective on the relations between the texts, from the level of single word to more complex multi-word lexicogrammatical sequences. I have argued that the consistency of the results with different variables suggests that what the analysis has pinpointed is variation at the level of formal structure rather than semantic content. In fact, as explained in Chapter 3 (Section 4.4), the MFWs and MFSs used in this chapter include all of the most frequent function words in the TTs, rather than low-frequency content words examined in the previous chapters.

Finally, the findings of Sections 3.1 and 3.2 have shown a marked similarity with the outcome of the previous chapters. In the context of the contemporary TTs, the similarity between Translations M and B, and their difference from Translation G mirrors the differences identified in Chapters 5 and 6, namely the different behaviours the translations display in the way they deal with the fictional representation of the African jungle and the African natives. This seems to suggest the existence of a coherent larger picture, a picture in which the various levels of scrutiny, although very different one from the other, nevertheless provide consistent results.

4. Comparing the TTs in relation to the ST

In the second part of this analysis, I will introduce the ST into the picture, using it as a baseline for the comparison of the TTs. This will allow me not only to examine the degree of similarity between the individual translations and the ST, but also to study

how the TTs compare to each other in relation to the original. However, given the language difference between the ST and the TTs, a direct comparison – as carried out in the previous section – is not possible. I cannot simply repeat the analysis in Section 3 adding the ST, because the MFWs in the original will be in English and, of course, they are not comparable with the MFWs in the Italian versions. Yet, an alternative type of comparison is still possible. Instead of comparing the frequency patterns of the MFWs in the whole texts, I will examine the behaviour of the MFWs in sections of the texts, and then compare the clustering of the sections across the texts. In more practical terms, I will:

- (i) divide HoD into 10 sections;
- (ii) examine how similar or different the sections are to each other based on the frequency patterns of the 250 MFWs in HoD;
- (iii) divide each TT into 10 sections;
- (iv) examine how similar or different the sections are to each other based on the frequency patterns of the 250 MFWs in each TT;
- (v) compare the overall clustering of the sections across the TTs and with the ST.

This method allows me to compare the TTs to the ST despite the language difference. However, this type of comparison is not equivalent to that performed in Section 3. In the previous section, I have compared how similar or dissimilar each translation is to the others based on the frequency patterns of shared MFWs. In contrast, in this section, I will not directly compare the patterns of shared MFWs across texts; rather, I will compare how sections behave across texts. I will show that the clustering of the sections in HoD is indicative of its content and structure. Comparing whether the sections in its translations cluster similarly to the original or not enables me to examine to which extent this internal content and structure of the ST is maintained in the TTs. This represents an alternative way to compare a text to its translation.

The same method is used by McKenna et al. (1999) to study Beckett's trilogy (*Molloy*, *Malone Dies*, and *The Unnamable*) and its French translations. They divide the STs into twenty successive segments and examine how these segments array themselves on the score plot, based on the frequency patterns of 99 MFWs. Then they

repeat the procedure with the French texts, so as to be able to compare the clustering of the sections across languages. Rybicki (2006) too employs a similar procedure in his analysis of three novels by Henryk Sienkiewicz and their two English translations. However, instead of using successive segments of texts as sections, Rybicki (2006) uses character dialogues. He examines how the idiolects of the major characters (represented by their dialogues) cluster on the resulting score plot, and then compare the clustering with the equivalent clustering of the dialogues in translation. In this way, Rybicki (2006) is able to discuss the nature of idiolectal differentiation in Sienkiewicz and then explore the effects of translation on it.

For the present analysis, I will follow McKenna et al.'s (1999) method and divide HoD and the TTs into successive segments of equal length. First, HoD is segmented into ten sections of 3,850 tokens each (apart from the last section which is slightly longer). Then, the TTs are divided into ten sections which correspond to the sections in the original. This means that their length differs to various extents to that of the ST sections (see Appendix 3 for a length comparison of the sections), but the English segments and the Italian ones can be said to be “matching” (cf. McKenna 1999: 152): they are aligned so as to start and end at the exact same points. Therefore, they are supposed to include the same content.

4.1 Section clustering in the ST

The first text investigated is HoD, which is used as baseline for the comparison of the TTs. Consider the resulting score plot in Figure 7.5. Figure 7.5 shows the overall degree of similarity of the 10 sections based on the frequency patterns of the 250 MFWs in HoD. As for the previous figures, the closer the sections are, the higher their degree of similarity.

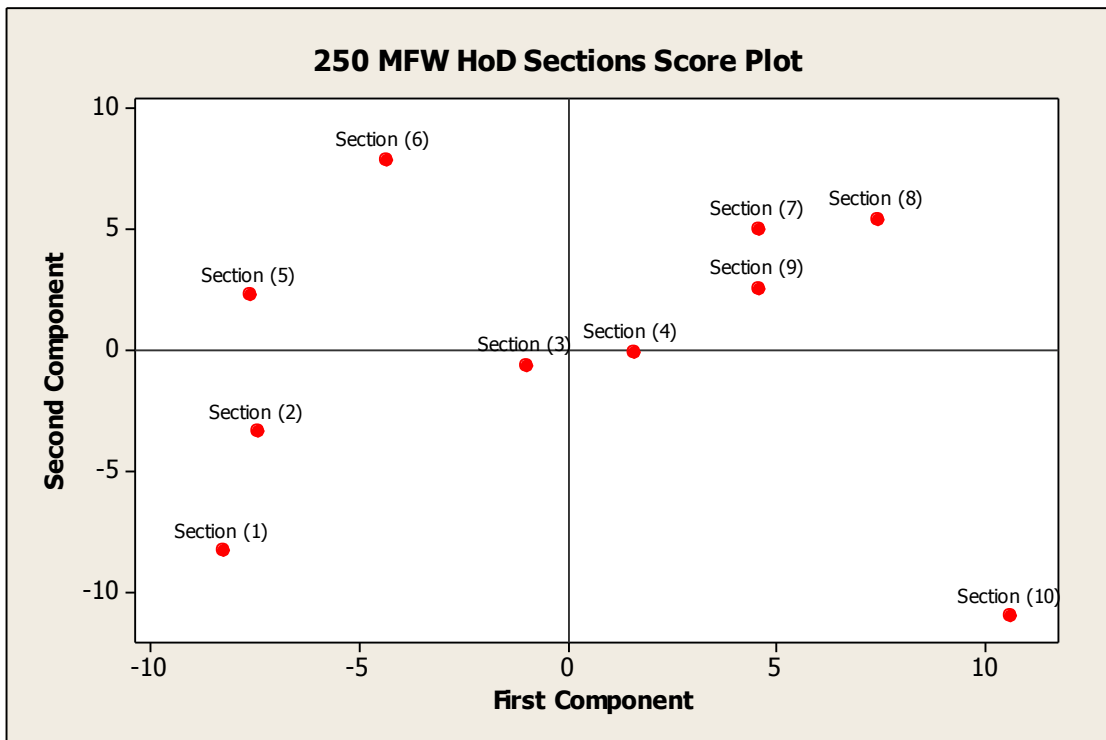


Figure 7.5 HoD sections score plot based on 250 MFWs

To make sense of this disposition, I have taken into account the content of each section to help to interpret their configuration. Sections (1) and (2) represent the introductory part of HoD set in London and in the continental city. The trip to Africa starts only in the final part of Section (2). In the score plot, these two sections cluster together. Section (10) is the other section not set in Africa: in it, Marlow goes back to Europe and meets Kurtz’s fiancée. Thus, the only other segment of HoD not set in Africa is independent and stands out in the bottom-right corner of the plot, far away from all the other sections. Although Sections (1) and (2), and Section (10) do not cluster together, the three of them appear in the bottom half of the plot, as opposed to all the other segments located in the upper half (apart from Section (3), which anyway is very close to the upper half). This marks the distinction between the sections set in Africa, the “African sections” (Sections (3) to (9)) and those set in Europe (Sections (1), (2) and (10)).

Given their different content and themes, it is possible to interpret why Sections (1), (2), and (10) do not cluster together, even though they are all set in Europe. Analysing Figure 7.6, the relative scatterplot of Figure 7.5, it is clear how the discriminators mark immediately the different functions of Sections (1) and (2), and

Section (10). Sections (1) and (2) establish the initial “who, where, and when” of the story, reflected by the presence of *men*, *river*, *old*, *big*, *sea*, and *work* as word-discriminators of these two sections. There are also many prepositions (*on*, *a*, *of*, *in*, *after* and *into*) among the discriminators of Sections (1) and (2), which contribute to set denotatively the story into its spatial and temporal coordinates. In contrast, the discriminators of Section (10) are of a very diverse nature. In the final part of HoD, Marlow is engulfed in the very heart of darkness, after his emotionally-intense experience in Africa. He has now to face Kurtz’s fiancée and recall back what he has experienced there, that “The horror! The horror!” (Conrad [1902] 1994: 100) he alone has been witness of. *Darkness*, *dark*, *death*, and *life* discriminate this section, together with verbs of mental activity as *know*, *knew*, *believe*, *understand*, *heard*, *seemed*, and the expected pronouns *she* and *her*. Ultimately, Sections (1) and (2) seem to be more descriptive in nature, as opposed to Section (10) that appears more reflective.

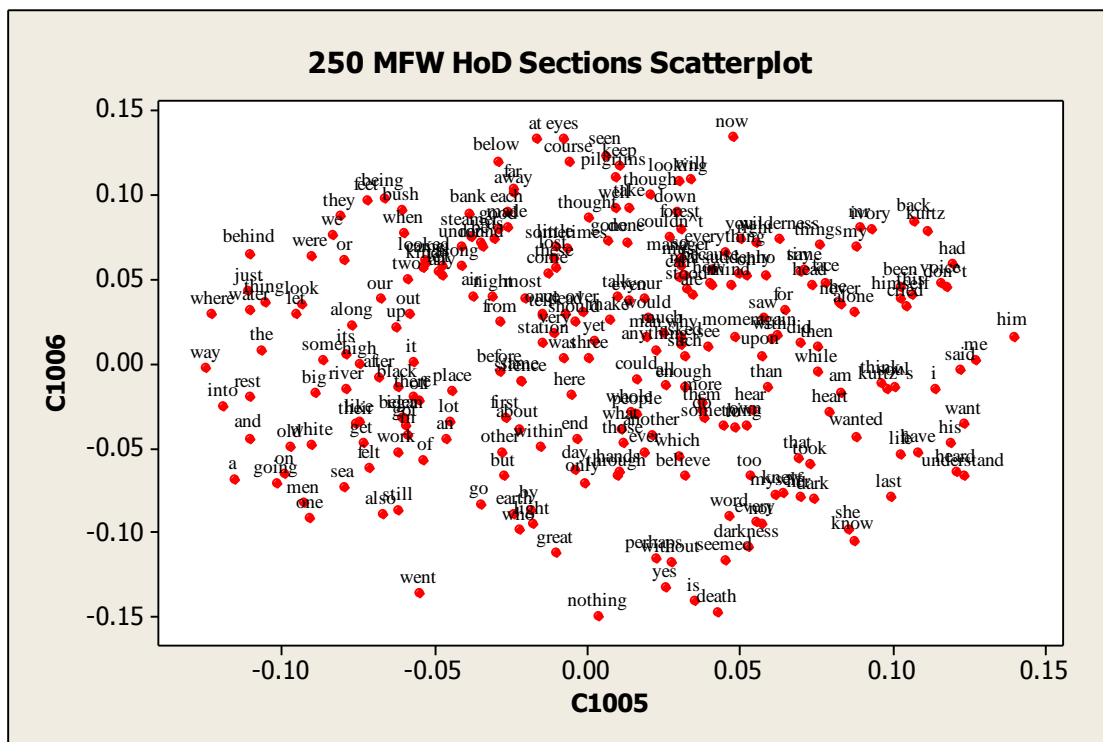


Figure 7.6 HoD sections scatterplot based on 250 MFWs

As for the upper half of Figure 7.5, the journey to Congo starts at the end of Section (2), while at the very beginning of Section (10) Marlow is already back in the continental city. So Sections (3) to (9) are those set in Africa. Words such as *pilgrims*,

station, manager, water, bush, bank, and steamer distinguish the top half from the bottom one. Looking more specifically at the top half, it is possible to identify smaller groups of sections. For example, Section (7), Section (8), and Section (9) cluster together in the top-right quadrant. This cluster is discriminated by the pronouns *he* and *himself*, the nouns *mind, voice, ivory, and face*, together with *mr* and *kurtz*. These last two discriminators are quite revealing: checking what these three sections have in common, it emerges that they share Kurtz as their focus. Marlow personally meets him in Section (8) but, since the beginning of Section (7), Kurtz more and more obsessively occupies the thoughts of Marlow and his dialogues with the people he meets in his journey towards Kurtz's station. Then, Kurtz dies at the end of Section (9). The difference of perspectives and style between Sections (7) to (9) and Section (10) justifies the gap between the former and the latter, even though all of them deal with a similar topic, namely Kurtz. In Sections (7) to (9), Kurtz is physically present in the narration, having a more active role than in Section (10), where he is just a memory of the past. Moreover, Sections (7) to (9) are more descriptive and concrete in their narration, while Section (10) is mainly dialogic.

4.2 Section clustering in the TTs

In this section, I will repeat the same procedure used to study HoD in the previous section, applying it to the analysis of the TTs. I will then compare the clustering of the sections in the TTs with the clustering of the sections observed in the original. It is important to point out that the absolute position of the sections in the score plots is not relevant in itself, rather what matters is the position of each section in relation to the others. It is normal, in fact, for these graphs to change orientation back and forth (Hoover 2012: online). For instance, if in Figure 7.5 Section (10) was located in the upper-left quadrant of the plot, it would not have affected the analysis as long as the distance of Section (10) from all the other sections was unaltered. Finally, what the comparison aims to identify is not the absolute position of each section, but their position in relation to the other sections.

Figure 7.7 below collects the score plots of the four Italian translations. I will first compare the plot of each TT with that of the ST, in Section 4.2.1. Then, in

Section 4.2.1, I will focus instead on comparing the clustering of the sections across the translations.

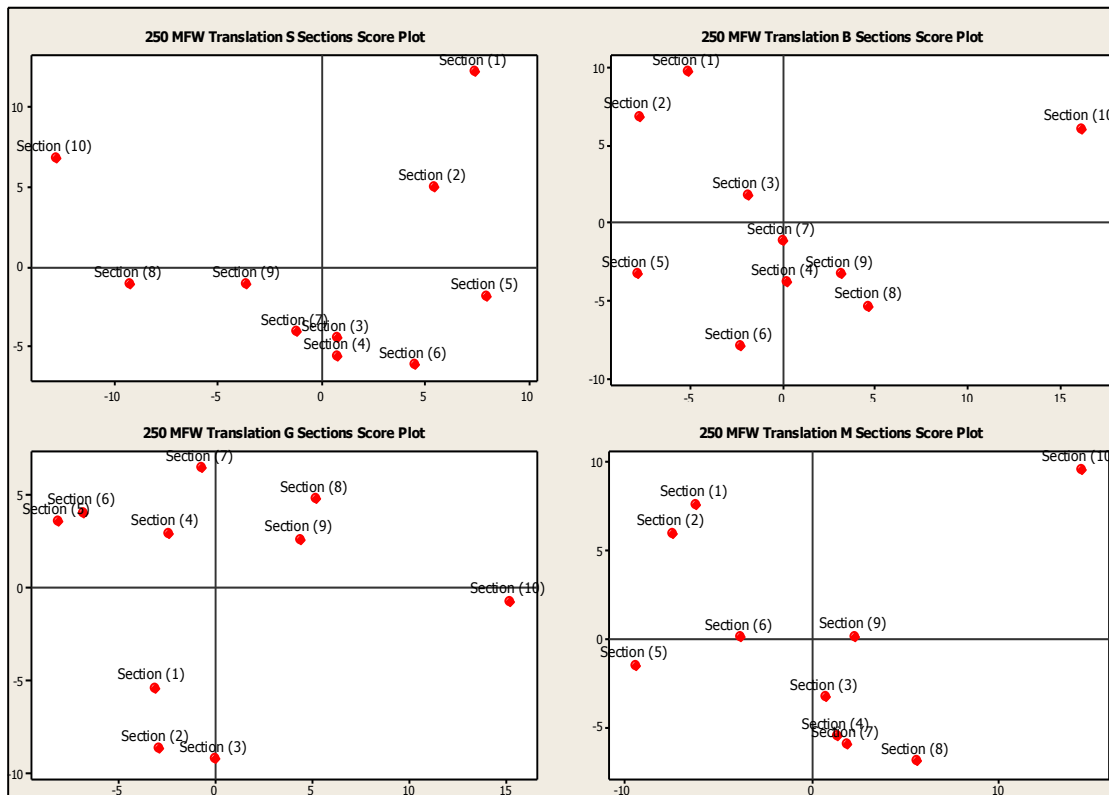


Figure 7.7 TT sections score plots based on 250 MFWs

4.2.1 Comparing the section clustering of the TTs to the ST

Starting from Translation 9 S, it may seem that the clustering is quite different from what has been shown in the original, but this is simply due to the orientation that has shifted. Actually, the overall disposition of the sections in relation to each other is rather similar between Translation S and the ST. First, Sections (1) and (2) are one next to the other in the same quadrant, although not as close as in Figure 7.5. Section (10) too occupies an equivalent position to that in Figure 7.5, and its relation with the other sections is maintained: it is separated from all the other sections, but the separation is not as clear-cut as in HoD, especially in relation to Section (8). Looking at the discriminators (see Appendix 2 for the scatterplots), there are some analogies with the relative sections in HoD, but this is far from being an absolute correspondence and the differences are more numerous than the similarities. In

particular, there are fewer prepositions around Sections (1) and (2), while around Section (10) what is missing are the verbs of mental activity.

It can also be noted that there is a more clear cut between the sections in the upper half and those in the bottom half. Those in the bottom half (the “African sections”) cluster more tightly than in Figure 7.5, even though Sections (7) to (9) do not cluster together as shown in Figure 7.5. Rather, Section (7) seems to be more similar to Sections (3) and (4). These changes are reflected in the scatterplot, which shows that, on the one hand, the general discriminators of the “African sections” are identifiable, but on the other, it is not possible to assign the specific focus on Kurtz to any cluster of “African sections”.

Translation B score plot has a shifted orientation too. Yet, the relation between Sections (1) and (2) is identifiable (stronger than in Translation S), as well as the position of Section (10) that is similar to what has been noted in the original. These three sections cluster separately from the “African sections”. However, Section (3) appears in the upper half of the plot, although it is closer to the “African sections” than to Sections (1), (2), and (10). Section (3) clusters closely with Sections (4), (7), (8), and (9). The scatterplot does not reveal any distinctive features that could explain this large cluster. The words that characterise the “African sections” are there, but the emphasis on Kurtz is missing. As far as the top half is concerned, the scatterplot shows that Sections (1), (2), and (10) are discriminated by similar words as in the original.

The score plot of Translation G has the same orientation as the score plot of HoD, so the two graphs are directly comparable. Section (3) moves down to cluster with Section (1) and (2), while Section (10) keeps its independent position. However, instead of being in the bottom-right corner of the quadrant, Section (10) in Translation G is located in the upper-right corner; this makes it much closer to the upper half of the graph, where the “African sections” are. Another difference compared to Figure 7.5 is that Sections (5) and (6) cluster closer together: their similarity is stronger than in the original. The opposite happens with Sections (7) to (9), which are further away from each other than in the original. Section (7) is closer to Section (4) than to Sections (8) and (9). The scatterplot shows an equivalent balance of similarities and differences compared to Figure 7.6. The discriminators are partially equivalent and

partially not, even in the case where the position of sections in the ST and in Translation G is analogous, such as in the case of Sections (1), (2), and (10).

Finally, the score plot of Translation M displays a shifted orientation, but similar clustering to what is shown in Figure 7.5 is still recognisable. Sections (1) and (2) are located one next to the other. Section (10) is in one of the corners of the plot, exactly as in the original, far away from all of the other sections. However, Sections (3) and (4) and Sections (7) and (8) mix together: Section (4) and Section (7) are almost overlapping, while Sections (3) and (8) are very close. The similarity between Sections (5) and (6) is stronger than in the original, even though they move up towards the middle of the graph, more than what happens in the ST graph. The analysis of the discriminators in the scatterplot shows a range of similarities and divergences compared to the original. Sections (1) and (2) are discriminated by similar words as in HoD, while Section (10) shares with the relative ST section just some of them, missing instead the verbs of mental activity. The bottom half of the graph lacks the focus on Kurtz, while it shows those general words that discriminate the “African sections” from the ones set in Europe.

Summing up, all the translations share the overall section disposition of the original, differing in some details but without any considerable alteration. The main features identified in the clustering of sections in HoD have been identified in the TT plots as well. First, all four of the TTs have Sections (1) and (2) close together in the same quarter of the graph. Second, Section (10) shows an independent behaviour and its position is isolated from the other sections in all of the TT score plots. Third, all of the TTs present the distinction between “African sections” and the sections set in Europe, although this division is more or less marked depending on the TT. It is in fact in the clustering of the “African sections” that it is possible to see variation between the TTs and the ST, especially in terms of the focus on Kurtz, which is not as recognisable as in the original. In the TTs, it is Sections (3), (4), and (7) that cluster close together, as opposed to Sections (7), (8), and (9). This cluster emerges among the “African sections” in Translations S, B, and M, but not in Translation G. This leads the discussion to the comparison of the TTs to each other.

4.2.2 Comparing the section clustering of the TTs to each other

Generally speaking, the four score plots show a similar arrangement of sections. This suggests that the translations are much more similar than different from each other, in this respect. They all show the distinction between the “African sections” and the sections set in Europe, the clustering between Sections (1) and (2), and the independent behaviour of Section (10). Although different from what seen in the original, they also show a similar cluster among the “African sections”, that is, Sections (3), (4), and (7).

However, as mentioned before, this cluster appears in all of the translations apart from Translation G. Overall, Translation G is the text that displays more independent behaviours compared to the other TTs. First of all, Section (3) clusters together with Sections (1) and (2). On the contrary, in the other TTs and in the ST, Section (3) groups with Sections (4) and (7), together with the other “African sections”. Moreover, in Translation G, Sections (5) and (6) almost overlap, while in all the other translations they are more distant from each other.

The difference of Translation G among the contemporary translations is a recurring outcome of this thesis. First of all, I have pointed out a similar result in Section 3.3, when discussing the overall degree of similarity of the TTs based on the frequency patterns of MFWs and MFSs. Moreover, in Chapters 5 and 6, I have shown that Translation G displays more differences compared to the other TTs, in relation to the way the fictional representation of the African jungle and the African natives have been reproduced. The present analysis has emphasised again the divergence of Translation G, showing that its difference from Translations M and B is also recognisable in the way the MFWs behave across the sections.

4.3 Different comparisons, different findings

The comparison proposed in this section has shown a different picture from the comparison carried out in Section 3. When the translations are compared in relation to the ST, I have argued that they seem to be more similar than diverse one from the other. However, when compared to each other, independently from the ST (as in Section 3 above), the translations are quite different. Translation S is very different

from Translations B, M, and G; Translation G is different from Translations B and M; only Translation B and Translation M seem very similar to each other. Yet, all four of the TTs show an equivalent section clustering as in the original (except from the differences identified in Translation G).

This distinction can be interpreted in light of the fact that, although the translations are four independent texts, they are in the end four versions of the same text. From a general point of view, the aim of fidelity to the original brings the translations close together, despite their characteristic features and idiosyncrasies. This suggestion is in agreement with Rybicki & Heydel's (2013) findings, in their study of the Polish collaborative translation of Virginia Woolf's *Night and Day*. They argue that texts by the same author cluster together even when they are translated by different translators. For example, two sections of Virginia Woolf's *Night and Day* by different translators are much more similar to each other than each translator-specific section with other translations by the same translator. The findings of Section 3 and of this section mirror Rybicki & Heydel's (2013) argument, as I have also shown that the original author's signal prevails in translation over the translator's one. This predominance could explain why Translation S too shares more similarities than differences with the other TTs when compared in respect to the ST, even though I have demonstrated that Translation S differs considerably from the other TTs, when the translations are compared to each other. Despite the fact that it has been written more than 60 years before the other translations, Translation S still presents the same clustering of the ten sections as the other TTs.

5. Stylistic overtones in translation

In the discussion above, I have shown that the section clustering is overall very similar in both the ST and the TTs, suggesting a high degree of similarity among the translations when they are compared to the original. This outcome may seem surprising at first, if one considers that the comparison is based on the frequency patterns of MFWs in different languages. There is in fact no one-to-one correspondence between the MFWs in the ST and those in the TTs. The 250 MFWs of the original are not a specular reflection of the 250 MFWs in the TTs, since the way

Italian is lexically and grammatically structured is very different from English. For example, the fact that Italian is a much more inflected language than English means that the frequency of very common function words, such as pronouns or articles, is divided across more word types than in English. This is because these function words have different inflected forms, agreeing with person, number, and gender. At the same time, Italian frequently omits personal pronouns as subjects: *I* ranks 4th in the MFW list of HoD, while the Italian counterpart *io* is only 37th. These are but few instances to illustrate how the divergences between the two languages result in MFW lists which are not one the direct translational equivalent of the other. Yet, the score plots show a very similar patterning across the texts, a patterning that imitates closely the one in the ST despite the interlanguage shift. Given these differences, how can we explain the similarities obtained?

First of all, it is important to mention that this outcome is in agreement with what other researchers have pointed out. McKenna et al. (1999) too, in their analysis of Beckett's Trilogy and its French translation, find that the sections of the STs and of the TTs pattern very similarly, despite the different words from different languages the analysis is based on. They wonder whether the fact that Beckett translated his own work in French could have affected their finding. However, Rybicki (2006) obtains similar results when comparing English translations of Polish novels, produced by professional translators as opposed to the original author. The patterning of the sections across texts are "mysteriously preserved" (Rybicki 2006: 102); they are so similar that the only discrepancies identified are attributed to the mismatch between the Polish MFWs and the English ones (Rybicki 2006: 102).

In order to suggest an explanation for this outcome, I will use McKenna et al.'s (1999) notion of 'stylistic overtones':

The network of relationships [among sections in the STs and TTs] we have described rests on linguistic structures that constitute the very basis of those "stylistic overtones" that readers depend upon in the formation of their literary judgments about texts. (McKenna et al. 1999: 169)

If we assume that all texts have their own style, we can hypothesise that each section of a given text also presents its own stylistic overtone, the result of the linguistic features that distinguish one section from the other. As sections are just an arbitrary division, it is not surprising that there are some that are very similar to others, since they might be parts of the same “chunk” of text, sharing the same stylistic overtone. Sections (7) to (9) in HoD, for example, may overlap in terms of stylistic overtones, as they represent an arbitrary segmentation in three chunks of a larger and unitary part where the homogeneous focus is Kurtz. At the same time, sections with a high degree of unlikeliness mirror the heterogeneity of style in the text: style changes, rises, flattens out, or sharpens in accordance with the communicative needs of the author, the narrator, and the narration. The stylistic overtone of these sections will be consequentially diverse. This is the case of Section (10), for instance, as it includes the most dialogic part of the short novel, without parallel in the rest of the text. Therefore, its stylistic overtone is markedly different, hence its distance from all the other sections. This hypothesis is in line with Egbert’s (2012) findings: in the investigation of the degree of variability within individual novels, Egbert (2012) identifies a high within-novel stylistic variation. In particular, he focuses on the variation of Dimension 1 (“thought presentation vs. description”) in two of Mark Twain’s novels, *The Prince and the Pauper* and *Tom Sawyer, Detective*. The score range of Dimension 1 in the two novels is smaller than Twain’s overall Dimension 1 range; yet, within the novels, the scores range drastically between one chapter and the others (Egbert 2012: 188), indicating considerable within-novel variation.

As suggested by McKenna et al. (1999), readers depend upon stylistic overtones to shape their literary judgments about texts. This can be equally true for sections of the text, and for translators too, as they do engage with the text as readers. Consequentially, it can be said that translators are affected by stylistic overtones in the formation of literary judgments about the various parts of a given text. This presupposes an attempt – conscious or not – on the part of the translator to reproduce the stylistic overtones of the original in translation.

Stylistic overtones, in their organisation, depend upon the frequency patterns of “linguistic structures”: these linguistic structures constitute the very basis of the overtones (McKenna et al. 1999: 169). In the context of the present analysis, the linguistic structures are represented by the frequency patterns of the MFWs. For

example, a high frequency of common prepositions may result in an abundance of prepositional phrases which generally mark a more descriptive or reflective tendency in writing (as discussed in Section 4.1); where they are sparser, on the other hand, it may be because the action is taking place upon a barer stage (cf. McKenna et al. 1999 for this and other examples). Similar MFW frequency patterns between ST and TTs result in similar stylistic overtones, whereas discrepancies in the patterns result in diverse stylistic overtones. An equivalent high frequency of prepositions in translation can therefore suggest that the TT maintains the descriptive style of that section in the original. However, this does not imply a direct translational equivalence between the MFWs; as said before, there is an interlanguage mismatch between them. The prepositions in the TT, for instance, do not necessarily have to be exactly the same one-to-one equivalent prepositions in the ST. What matches instead is the linguistic organisation of the ST and TTs, of their different sections, mirrored in the use of different but equivalent MFWs: functional equivalence, rather than word equivalence (see Chapters 1 and 3).

I have explained in Chapters 1 and 3 and shown in Chapters 5 and 6 that words develop their meaning through the network of relationships they establish with each other. The MFWs have been seen as the markers of these relationships (McKenna et al. 1999: 152). Changing the markers alters the relationships, hence meaning is affected. Keeping them unaltered preserves the network, hence meaning is unaffected. A translator who wants to reproduce similar stylistic overtone as in the ST (or in sections of the ST) would try to replicate the network of word relationships, although this assumes the use of language-specific markers. Alternatively, a diverse stylistic overtone can result from the alteration of the word network, due to a shift in the use of the markers. The alteration of MFW frequency patterns in translation, then, contributes to give a different “stylistic flavour” to the extract.

In the light of this, it is possible to account for the similarities identified between sections in the ST and TTs as an attempt on part of the translators to maintain the ST stylistic overtones, reproducing an equivalent linguistic organisation but with language-specific tools. The functional reproduction of equivalent stylistic overtones throughout the whole translation can be seen as what makes translations mainly similar to the original when they are compared as entire texts. However, I have shown in the previous two chapters that when the analysis focuses instead on specific textual

features, then it is possible to identify discrepancies between translations. These two levels are mutually related: translations are mostly similar to the original because they reproduce with functionally equivalent structures the overall stylistic overtones of the ST; however, the alterations that translation always implies can affect fundamentally the reception of the TT, no matter how minor these alterations may seem in the overall picture of the whole text.

6. Conclusion

In this chapter, I have compared the translations as whole texts, using PCA to provide a statistical measure of their overall variation, based on the frequency patterns of items which have not been taken into account in the previous chapters. The occurrence of these items is so common that it is unlikely they are consciously regulated by authors: they have been in fact seen as representation of “authorial wordprints” (Hoover 2002) or “stylistic signature” (Burrows 2002a) and, as such, considered author- and text-specific.

This chapter has shown the overall relations among the TTs, though with some limitations. The analysis has demonstrated that Translation S is very different from the contemporary TTs; that Translation G is the least similar among the contemporary translations; finally, that Translations M and B are very similar to each other. However, the study of these differences has been affected by the degree of similarity among the translations. The more different the TTs are, the more clearly their differences have been pointed out; the more similar the TTs are, the more difficult it is to distinguish between one TT and the other. For example, in the case of Translation S and the contemporary TTs, the analysis has shown clearly why they differ. On the other hand, with Translations B, M, and G, the differences were less striking, so it has been more difficult to indicate the reasons behind their variation.

Despite these limitations, I have shown that the findings of this analysis are in line with the outcome of the previous analyses. Two key points have independently emerged in Chapter 5, Chapter 6, and this chapter: the fact that Translation S distinguishes itself for belonging to a different temporal and socio-cultural context, and the fact that Translation G is the most distinctive among the contemporary TTs.

The two perspectives that these chapters represent, then, match: the whole-text comparison and the feature-specific comparisons are in agreement. This coherent picture suggests an interaction of the textual levels accounted for in each chapter.

The examination of the lexical level has taken centre stage in the whole thesis, because of the very nature of the approach adopted in this work. Yet, even with this shared feature, Chapters 5 and 6 and this chapter have presented different perspectives. In Chapters 5 and 6, I have focused on the semantic level. Semantic preferences and semantic prosodies have played a fundamental role, and the items examined were mostly content words. In this chapter, however, I have focused on high-frequency items, mainly function words. Function words have been considered the connectors between content words, so their study can emphasise the structures they establish between content words. I have argued that the difference in their frequency patterns across texts can indicate divergences at the level of sentence and text construction. This has been supported by the consistency in the results of the analyses using MFWs, two-word MFSs, and three-word MFSs as variables: these are three successive steps in the same process of sentence construction. The TTs display a preference for given grammatical classes/categories, as shown by the clustering together of several forms of the same item in correspondence of a specific translation. It is the way translators structured their sentences, their texts, rather than differences at the level of content, which accounts for the divergences pointed out by PCA. The two levels – the content level and the structure level – match coherently, confirming once more their interplay, their being *de facto* the two facets of the same phenomenon.

From a translational point of view, this chapter has emphasised that, when the TTs are compared to the ST, similarities prevail over divergences. I have argued that it is the aim to be faithful to the ST that moves the TTs' differences to the background, so that the ST's linguistic signal prevails over the individual translators' ones. Unless motivated by a different agenda, translators try to be as faithful as possible to the original, using the tools that the TL provides them with. The individual translators' idiosyncrasies do not seem to overcome the ultimate aim of faithfulness to the ST. Translation G is the TT that has shown more variation in terms of section clustering but, even in this case, the degree of similarity to the ST is higher than the degree of difference.

I have argued that the overall similarity of the TTs to the ST is the result of the translator's attempt to reproduce similar stylistic overtones throughout the text. Stylistic overtones rest upon the network of relationships that linguistic structures establish (McKenna et al. 1999: 169), and these linguistic structures depend on the use of function words, the same function words I have used to carry out PCA. Equivalence in the usage of function words contributes to recreate the same stylistic overtone. However, the equivalence is functional, rather than based on word correspondence. The translators aim to reproduce the different stylistic overtones of the ST adopting functionally equivalent and language-specific tools: this is why similarities in the clustering are still identifiable despite the differences in the MFW lists used.

CHAPTER EIGHT

Concluding discussion

1. Introduction

This chapter will provide a concluding discussion of the contributions and implications of the present study. In Section 2, I will discuss the major findings that have emerged from this thesis. In Section 3, I will put forward methodological considerations focused on the different levels of comparison that the previous chapters have presented. In Section 4, I will discuss the limitations of this study, while in Section 5 I will suggest some directions for expanding this research further.

2. Major findings

Three major findings have emerged from the analyses in the preceding chapters: the role of textual patterns as building blocks of the fictional world; the effect of translation on the relation between textual patterns and the fictional world; and the link between the comparison at the level of individual features and that at the level of whole texts. I will discuss each of these points individually, in Sections 2.1, 2.2, and 2.3 respectively.

2.1 Textual patterns as building blocks of the fictional world

In this thesis, I have argued that textual patterns can function as building blocks of fictional worlds in literature. My study has foregrounded the role of specific patterns in constructing the fictional representation of the African jungle and of the African natives, both quantitatively and qualitatively. This has been achieved in Chapter 4, Chapter 5, and Chapter 6, addressing Research Questions 1 and 2.

In corpus linguistics, KW analysis is frequently used to commence a study by identifying a pool of items to examine in detail, items that are characteristic of the

text/corpus under investigation and easily manageable in terms of quantity. This thesis is not an exception; however, I have shown that KW analysis can also be employed to establish a connection between the lexical level of a text and its themes. Generally speaking, KWs are considered to be particularly effective in signalling the topics of a text – in other words, what a text is about (Scott 2015a: online). However, in his study of HoD, Stubbs (2005: 11) argues that KWs can only indicate superficial topics, while underlying themes cannot be recognised. He takes *river* as an example, claiming that this word indicates a superficial topic but not a theme: “this is not a book ‘about’ a river” (Stubbs 2005: 11). This is certainly true, but it is also a partial perspective. Although individual KWs can hardly account for a theme in a book as complex as HoD, I have shown that groups of KWs offer a different picture. Groups of semantically and thematically related words are able to establish a link between a given theme and the lexical level of the text.

I have argued that, to establish this link, the simple identification of “aboutness” KWs (Scott 2015a: online) is not enough. A more fine-grained discrimination is necessary, a categorisation of KWs that takes into account their potential to build the fictional world and convey thematic concerns. This potential is brought out by grouping KWs into fictional world signals and thematic signals (Mahlberg & McIntyre 2011). I have demonstrated that groups of related KWs are able to establish a link with specific aspects of the text, such as its characters, its setting, its atmosphere, and its themes. The resulting “Africa words” and “African words” include fictional world signal KWs, thematic signal KWs, and thematically related words: the study of the “Africa/African words” represents a well-grounded way to examine how major themes in HoD are connected to the linguistic level of the text.

The patterns that the “Africa words” and the “African words” establish function as building blocks of the fictional world, because they linguistically construct the representation of the jungle and of the natives in the text. This analysis has identified three main patterns associated with the items in the two groups: lexical repetition, semantic preferences, and semantic prosody. Their persistent co-occurrence shapes the depiction of these fictional entities, defining them as discussed in Chapter 5 (Sections 2 and 3) and Chapter 6 (Section 2), and as summarised in Table 8.1 below.

Table 8.1 Summary of the representation of the jungle and of the natives

African jungle	African natives
Personified	Dehumanised
Metaphorical representation	Physical representation
Shapeless place	Represented as groups rather than individuals
Impenetrable place	Incomprehensible to the colonisers
Dark place	Related to negative states or situations

The consistency of these patterns ensures that every time the reader comes across one of the “Africa words” or “African words”, he/she finds them associated with the features summarised in Table 8.1. Occurrence after occurrence, these features literally build up the representation of the fictional world throughout the text – the patterns are indeed building blocks.

The implications of this specific textual representation go beyond the construction of the fictional world. The fictional world, in turn, is related to the text’s literary themes. Throughout Chapters 2, 4, 5, and 6, I have shown how the depiction of the African jungle relates to the theme of Africa and its representation, while the depiction of the African natives is linked to the theme of race and racism. As explained in Chapter 2, a postcolonial reading of HoD emphasises the way Africa and Africans are represented because such representations can be regarded as indicative of Conrad’s ideologies and attitudes towards the colonial enterprise. Africa is in a dialogic opposition to the Western world, and through this dialogue issues of imperialistic power and domination can be analysed. Moreover, given its popularity, HoD has contributed significantly to shape the reception of Africa, both in literature and in popular culture. Similarly, Achebe’s (1990) criticism, and the critical debate that he has generated, considers the way Africans are represented in HoD as central in the discussion about race and racism. Therefore, when we look at how Africa and Africans are described in HoD, we are not only looking into the settings and characters of the novel, but we are also considering how these aspects of the fictional world reflect more complex issues.

Given the close connection between the fictional world and major themes of HoD, I would argue that the patterns that this thesis has identified have a double role. Not only do they function as building blocks of the fictional world, but they also contribute to the two major themes. I have shown that the persistent use of repeated vocabulary and semantic fields throughout the text establishes a lexical network

among each item in each group of words (the “Africa words” and the “African words”) that makes the fictional representation of the jungle and of the natives cohesive. This notion of the cohesive network can be also applied to the themes: the consistent use of the same fictional representation throughout the text establishes a thematic network among each word in each group that contributes to the overall process of theme construction. In a sense, the textual patterns are building blocks of both the fictional world and of the major themes.

This twofold function reflects the two perspectives that this thesis has sought to take into account: the text-based and the reader-based perspectives. On the one hand, I have provided insight into the distinctive linguistic features that characterise HoD as an artistic work, compared to other texts. The analyses in Chapters 5 and 6, for example, build on the outcome of the KW analysis in Chapter 4: KWs are automatically identified on the basis of their unusual frequency. Examining them has enabled me to provide insight into the intrinsic linguistic features of HoD. On the other hand, these text-based features have been discussed in light of reader-based perspectives. Whereas KWs are statistically defined on the basis of textual features – their frequency compared to a reference corpus –, the categories I have used to group them reflect instead how the critics understand the text, thus following criteria that are not necessarily text-based. The “Africa words” and “African words” have emerged as relevant items for the study of HoD’s major themes; these themes, in turn, mirror a shared and consolidated interpretation of the text, that of the literary criticism. The conjunction of these two perspectives is in line with the very nature and aim of corpus stylistics: relating linguistic description to literary appreciation.

My analysis has shown that discursive constructions can also be identified and studied in literature. Corpus methods have frequently been employed in discourse analysis to analyse the linguistic representation of different social and cultural groups, such as refugees and asylum seekers (Gabrielatos & Baker 2008), suffragettes and suffragists (Gupta 2013), and Muslims (Baker et al. 2013). These studies have shown that corpus approaches can uncover ideologies and stances as they are reflected by discursive representations (Gabrielatos & Baker 2008: 6). However, research in this field is mainly focused on language in the media. I have demonstrated that a similar approach can be applied in literary studies, because all literary texts are tied to a particular social context. Analysing the linguistic representation of characters or

groups in a fictional text thus provides insight into larger social discourses. This is the case for the representation of Africa and Africans in HoD that, although specific to a given text, establishes links with a wider historical context.

2.2 Textual patterns and the fictional world in translation

The comparison between the ST and the TTs has shown the effects of translation on the relation between textual patterns and the fictional world. The linguistic alterations that affect the patterns in turn affect the representation of the African jungle and of the Africans in the Italian versions. Chapters 5 and 6 have provided evidence for this point, addressing Research Question 3.

Texts are organised on the basis of word relations. In Chapter 3, I have argued that meaning is created through the relationships between words, rather than emerging from their use in isolation (Sinclair 1991, 2004). Similarly, text cohesion is established through grammatical and lexical networks (Halliday & Hasan 1976). In Chapters 5 and 6, I have shown that these relations and networks take centre stage in shaping how the jungle and the natives are depicted. Each occurrence of the “Africa words” and “African words” relates to the other occurrences through a lexical chain of repeated vocabulary and semantic fields. This network establishes a cohesive representation of the jungle and of the natives. My analysis has emphasised that the discrepancies produced by translating are not simply the result of shifts at the level of the individual word, but are rather the consequences of the individual shifts on the word relations.

In both Chapter 5 and Chapter 6, I have shown how the translators maintain the semantic preferences identified in the original, reproducing equivalent semantic fields. However, the relation between the fields and the “Africa/African words” is altered. Translating *wilderness* and *darkness* with multiple terms significantly affects the network existing between the fictional entity and its textual representation. Similarly, replacing all the uses of *nigger/niggers* and *negro* with *nero/neri* (“black”/“blacks”) creates a mismatch between the way black people are described (semantic fields) and the way they are referred to (“African words”), a mismatch that

is not present in the original. Thus, discrepancies in the word relations result in discrepancies in the fictional representations.

Given the existing link between linguistic construction and literary themes, I have argued that alterations to the fictional world have the potential to affect how these themes are perceived in the target culture. If the description of Africa and black people is key to our understanding of how HoD engages with issues of colonial power and race, then it is not surprising that the way we perceive these issues can change if the descriptions of Africa and Africans themselves are altered. My analysis has foregrounded how alterations to the micro-level of individual textual features can affect the macro-level of the text, its interpretational implications (Munday 2008). In this respect, such alterations can be said to have manipulative effects.

In Chapter 1, I have discussed the concept of translation manipulation, outlining the criteria in accordance to which it is usually defined in translation studies research. My analyses in Chapters 5 and 6 have suggested a rethinking of the phenomenon, as its usual definition may be too restrictive, particularly the view that manipulation is always intentional and that it always has an ideological motivation. I have agreed with Dukāte (2009) in recognising the existence of both intentional and unintentional manipulation. More importantly, I have supported this view by providing a comparison of two different types of shifts, namely obligatory and non-obligatory (Toury 2004). In Chapter 5, I have discussed obligatory shifts, resulting from the lack of a direct equivalent: *wilderness* does not have an equivalent in Italian, while *darkness* has several equivalents, depending on the context. In Chapter 6, I have examined non-obligatory shifts, i.e. those which are not due to a lack of an equivalent, but rather to a choice of the translator. I have suggested, for example, that translating *nigger* with *nero* (“black”) or *black* with *negro* (“nigger”) are non-obligatory shifts, given that both ST terms have a direct Italian equivalent. My study has shown that the two different shifts can have equally manipulative effects on the TT, independent of whether they were intentional or unintentional. From an effect-oriented perspective, both shifts have the potential to produce the effect that is usually attributed to manipulation: misleading the reader and distorting the original text.

2.3 Different levels of comparison

The ST and the TTs have been compared in two different ways: at the level of individual features and at the level of whole texts. Whereas the first comparison has shown that the TTs differ considerably in the way they reproduce the individual features analysed, the second comparison has instead revealed that the translations are more similar to the ST than different. This perspective has been put forward in Chapter 7, providing an answer to Research Question 4.

The outcome of PCA improves our understanding of the relation that a single ST can establish with multiple TTs. It tells us that, in translation practice, the original text's signal prevails over the translation's signal. I have suggested that the individual idiosyncrasies of the TTs are not strong enough to overcome the faithfulness to the ST: despite the individual differences, the four TTs are very similar to each other when compared to HoD. This finding supports previous research on multiple translators working on the same text (Rybicki & Heydel 2013), which has shown that texts by the same author that are translated by different translators display less variation among them than texts by the same translator but by different original authors.

I have used the concept of 'stylistic overtones' (McKenna et al. 1999) to propose an explanation for this finding, an explanation that takes into account the way PCA is computed. PCA provides a statistical measure of the degree of similarity between texts, based on the frequency patterns of the MFWs in those texts. The MFWs in any text are mainly function words. These items can be considered as the markers of the relationships between content words and, as such, they also mark everything these relationships entail (McKenna et al. 1999: 152). For example, a high frequency of prepositions usually indicates a correspondent abundance of prepositional phrases that, in turn, can signal a descriptive tendency in writing. Similarly, frequent use of first and second person pronouns can indicate a dialogic passage. These fundamental linguistic structures contribute to the overall stylistic overtone of each part of a text. By reproducing equivalent linguistic structures, the translators maintain the same stylistic overtones of the original throughout the TT. As a result, the translations are similar to the ST, because they keep the overall stylistic structure of the original unaltered.

This result allows me to consider the findings of Chapter 5 and Chapter 6 through the wider perspective of the general relationships between the texts. Even though the TTs display marked discrepancies at the level of specific content words, these alterations from the ST do not seem to significantly affect the stylistic overtones of the original. However, further research is needed to establish the effect of these local discrepancies on the interpretation of the text. For example, Gabrielatos & Baker (2008: 21) suggest that semantic preferences and prosodies can have a central role in the way a reader perceives a text. While reading, we do not store a verbatim version of the text in our long-term memory, but rather its understanding and interpretation (Clark & Clark 1977: 138, 141). Gabrielatos & Baker (2008: 20) propose that semantic preferences and prosodies, by virtue of the frequency of related words and semantic fields, can trigger associations that are then maintained in the understanding of the text. Therefore, the question would be whether the local discrepancies at the level of specific semantic preferences/prosodies have an effect on the overall interpretation of the text, despite the fact that they do not affect significantly the general similarity between the TT and the ST. I will discuss this avenue of further research in Section 5.

3. Methodological considerations

The methodological approach to the study of literary translation in this thesis is based on the comparison between the ST and the TT. In particular, I have suggested a method that establishes a comparison from two distinct but interrelated perspectives. In Chapters 5 and 6, comparison has been established on the basis of individual linguistic features, whereas in Chapter 7 HoD and its translations have been compared as whole texts. These two different approaches have shed light on different aspects of the texts: the semantic level of content words and the structural level of function words, respectively. Yet, despite the diverse approaches and items taken into account, the two methodological perspectives match in their results.

Chapters 5 and 6 have focused on how the patterns of low-frequency, yet thematically relevant, content words have been reproduced in translation. Both analyses have shown that Translation G is distinct from the other TTs, while the other

translations share similar tendencies. Specifically, Translation G is the TT least affected by the loss of lexical cohesion, as it uses only two different terms to translate *wilderness*. It is also the only translation that consistently mitigates the potential racist implications of the ST's lexical choices. In contrast, Translations M, B, and S use a large range of items for *wilderness*, and maintain or intensify the potential racist implications of the original. These differences can be seen as more marked within the contemporary translations, given that they are products of the same socio-cultural context, as opposed to Translation S.

A similar picture has been obtained in Chapter 7. With PCA, the comparison of the TTs has been established on a wider ground, covering the majority of the tokens in the texts, including all of the most frequent function words. On this basis too, the same relations are recognisable: Translation G is the most different TT among the contemporary ones; Translations B and M are very similar to each other; Translation S stands out for its diachronic distance. Thus, the two levels of comparison support each other, leading to the same unvarying picture.

This convergent picture conforms to the view of a text as a complex unit in which different constituents interact cooperatively in order to construct a cohesive whole. The compatibility of the results confirms that the two comparisons – one focused on the content level, based on low-frequency content words; the other focused on the structural level, based on high-frequency function words – do not represent independent views, but rather different, yet interacting, textual levels. Finally, the convergence of the comparisons indicates that the methodology provides a valid picture of the relations between the texts, a picture that is simultaneously multi-layered and coherent.

4. Limitations of this study

This thesis has used HoD and its Italian translations in order to study the role of textual patterns as building blocks of the fictional world and the effect of translation on this role. However, this is by no means exhaustive, for HoD cannot be considered representative of literature in general, nor can the relation between this specific ST and its TTs be seen as indicative of all translation phenomena. Therefore, the

contributions of this thesis need to be considered in the light of this limitation. My study has built on previous research and my results have been corroborated by extant studies; this supports the generalisability of my contributions. However, to broaden the claims to literature and translation in general, more texts and translations need to be taken into consideration.

Additionally, this study could not compare the features in the TTs against a reference corpus of Italian translations. A corpus of novels translated into contemporary Italian is not available, and building one large enough to function as a reference corpus was beyond the scope of this project, especially given copyright considerations. Nevertheless, such a reference corpus could have been useful to expand the evidential base provided by this thesis. For example, it would be possible to check whether marked features in the ST are equally marked in the TTs, or whether translators standardised the feature replacing it with a more common one in the (translated) TL. However, it is worth pointing out that this research has aimed to focus on the original author's style in translation, as opposed to the translators' style. This aim has required the use of reference corpora only to study the ST and identify HoD's stylistic features. Once these features have been identified, the TTs have been examined on the basis of the analysis of the original. Thus, the use of a reference corpus of novels translated into contemporary Italian would be most useful to specifically investigate the style of the translators. I will return to this suggestion in the next section.

5. Further research

As suggested in Section 3, a potential avenue to develop this research further would be to employ reader-response analysis. My thesis has aimed to examine the relation between textual patterns and the fictional world, and to study the effect of translation on this relation. I have argued that some aspects of the fictional world are connected to major literary themes, and have suggested that the translation alterations to the patterns have the potential to trigger discrepancies in the way the themes themselves are perceived by the target reader. This suggestion could be tested by drawing on data

collections for reader-response analysis; questionnaires and interviews with readers would provide direct insight into the effects of linguistic alterations on the reader.

The response of SL readers to a given extract from the ST can be compared to the TL readers' response to the same extract in the TT, so as to test whether the translator's alterations manipulate the perception of the extract. Specifically, the reader's perception of the fictional representation of Africa and of the natives in the ST can be compared with the reader's perception of the same elements in the TTs. If the reader was provided with a series of extracts focusing on the wilderness, would he/she perceive *wilderness* as a cohesive fictional entity with a defined textual representation in the ST? In turn, would a TL reader miss the cohesive network in the TT as a result of the multiple terms used to translate *wilderness*? Moreover, would the reader's response to a translation where the natives are described mainly in terms of *negro/negri* ("nigger"/"niggers") be different from the response to a translation where all the potentially racist terms have been replaced by *nero/neri* ("black"/"blacks")? If so, would this difference have an effect on the perception of the other aspects of the fictional representation, i.e. the semantic preferences and prosody?

In addition to offline studies involving interviews and questionnaires, online psycholinguistic experiments could be carried out. For example, Mahlberg et al. (2014) use eye-tracking to investigate how body part clusters are read in Dickens's work. They relate reading times to readers' awareness of body part clusters, aiming to understand what effects such clusters can have on the understanding of characters. Similarly, eye-tracking could be used to study the reading times of the patterns that are related to the fictional representation of the jungle and of the Africans. Would the reading times of the same passages change between the ST and the TT, as a consequence of the translators' alterations? The development of this line of enquiry would contribute to the study of the effects of manipulation on the target reader in a more empirical fashion, allowing the researcher to relate specific textual alterations with potential interpretative implications.

An additional proposal for further development would be to shift the focus of the analysis from the ST to the TTs. In Chapters 1 and 3, I have explained that my thesis focuses on the study of Conrad's style in translation, rather than on the translators' style. Building on the analysis of the original's style carried out here, this study can be extended further by considering the style of the individual translators.

Baker (2000) maintains that if we wish to argue convincingly that translation is not simply a reproductive activity, but that it also involves creativity, then we need to pay more attention to the style of the translators and study it independently from the ST. Excluding the ST completely, however, could be counterproductive in some cases, because it is sometimes virtually impossible to separate out neatly the translator's own stylistic features from the intrinsic features of the ST. Huang & Chu (2014) and Huang (2015) in fact suggest to take both perspectives into account, as at times a feature of the TT can be the result of both the translator's individual style and the influence of the ST/SL. My study has already offered insights into the style of HoD and how the translators dealt with it; therefore, this further development of the present research would benefit from my findings to improve the understanding of the nature of the translator's style. Combining the two perspectives – the study of the original author's style in translation and the study of the translators' style – would provide a more nuanced picture of the relation between style and translation.

To undertake this advanced line of research, however, additional reference corpora would be needed. Both Baker (2000) and Saldanha (2011a) explain that, in order to define a linguistic feature as a translator's style, such a feature should be identifiable across a range of TTs by the same translator and distinguishable from the style of other translators. In this specific case, reference corpora of Italian translations would therefore be necessary. At the moment, these corpora are not available and building them would not be easy due to copyright restrictions. If such corpora were available, it would be possible to compare the TTs on the basis of their own specific features, as opposed to how the TTs rendered a specific feature of the ST. For example, carrying out a KW analysis of the TTs would emphasise those items that characterise each translation. These items can then be compared to the original, in order to examine whether they have been triggered by ST features or whether they represent actual stylistic idiosyncrasies of the translators.

REFERENCES

- Achebe, C. (1990). An image of Africa: Racism in Conrad's *Heart of Darkness*. In C. Achebe, *Hopes and Impediments: Selected Essays* (pp. 1-20). New York, NY: Anchor Books. (Original work published 1977)
- Ajello, N. (2008, January 27). La guerra dei pronomi nell'Italia in orbace. *La Repubblica*. Retrieved from <http://ricerca.repubblica.it/repubblica/archivio/repubblica/2008/01/27/la-guerra-dei-pronomi-nell-italia-in.html>
- Allington, D. (2006). First steps towards a rhetorical psychology of literary interpretation. *Journal of Literary Semantics*, 35(2), 123-144.
- Bachmann, I. (2011). Civil partnership – “gay marriage in all but name”: A corpus-driven analysis of discourses of same-sex relationships in the UK Parliament. *Corpora*, 6(1), 77-105.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233-250). Amsterdam, Netherlands: John Benjamins.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Someres (Ed.), *Terminology, LSP and Translation Studies in Language Engineering: In Honour of Juan C. Sager* (pp. 175-186). Amsterdam, Netherlands: John Benjamins.
- Baker, M. (1999). The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics*, 4(2), 281-298.
- Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target*, 12(2), 241-266.
- Baker, M. (2011). *In Other Words: A Coursebook on Translation* (2nd ed.). London, UK: Routledge.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London, UK: Continuum.
- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh, UK: Edinburgh University Press.

- Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourse of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-306.
- Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus-driven analysis of representations around the word 'Muslim' in the British Press 1998-2009. *Applied Linguistics*, 34(3), 255-278.
- Barlow, M. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, 18(4), 443-478.
- Baroni, M., & Bernardini, S. (2003). A preliminary analysis of collocational differences in monolingual comparable corpora. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference* (pp. 1-10). Lancaster, UK: UCREL, Lancaster University.
- Bassnett, S. (2007). Culture and translation. In P. Kuhiwczak & K. Littau (Eds.), *A Companion to Translation Studies* (pp. 13-23). Clevedon, UK: Multilingual Matters LTD.
- Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4(2), 119-139.
- Bernardini, S. (2007). Collocations in translated language: Combining parallel, comparable and reference corpora. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson, *Proceedings of the Corpus Linguistics 2007 Conference* (pp. 1-16). Birmingham, UK: University of Birmingham.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge, UK: Cambridge University Press.
- Binongo, J. N. G., & Smith, M. W. A. (1999). The application of principal component analysis to stylometry. *Literary and Stylistic Computing*, 14(4), 445-465.
- Blum-Kulka. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17-36). Tübingen, Germany: Gunter Narr Verlag Tübingen.
- Boase-Beier, J. (2004). Saying what someone else meant: Style, relevance and translation. *International Journal of Applied Linguistics*, 14(2), 276-287.
- Bosseaux, C. (2004). Translating point of view: A corpus-based study. *Language*

- Matters: Studies in the Language of Africa*, 35(1), 259-274.
- Bosseaux, C. (2006). Who's afraid of Virginia's *you*: A corpus-based study of the French translations of *The Waves*. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 51(3), 599-610.
- Brantlinger, P. (1985). "Heart of Darkness": Anti-imperialism, racism, or impressionism? *Criticism*, 27(4), 363-385.
- Brooker, P., & Thacker, A. (2005a). Introduction: Locating the modern. In P. Brooker & A. Thacker (Eds.), *Geographies of Modernism: Literatures, Cultures, Spaces* (pp. 1-5). London, UK: Routledge.
- Brooker, P., & Thacker, A. (Eds.) (2005b). *Geographies of Modernism: Literatures, Cultures, Spaces*. London, UK: Routledge.
- Burrows, J. (1987). *Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method*. Oxford, UK: Clarendon.
- Burrows, J. (2002a). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267-287.
- Burrows, J. (2002b). The Englishing of Juvenal: Computational stylistics and translated texts. *Style*, 36(4), 677-750.
- Carter, R. (2004). *Language and Creativity. The Art of Common Talk*. London, UK: Routledge.
- Catford, J. C. (1965). *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford, UK: Oxford University Press.
- Čermáková, A. (2015). Repetition in John Irving's novel *A Widow for One Year*: A corpus stylistics approach to literary translation. *International Journal of Corpus Linguistics*, 20(3), 355-377.
- Čermáková, A., & Fárová, L. (2010). Keywords in Harry Potter and their Czech and Finnish translation equivalents. In F. Čermák, A. Klégr, & P. Corness (Eds.), *InterCorp: Exploring a Multilingual Corpus. Studie z korpusové lingvistiky, svazek 13* (pp. 177-188). Praha, Czech Republic: Nakladatelství Lidové noviny/Ústav českého národního korpusu.
- Cheng, W. (2009). Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. In J. Flowerdew & M. Mahlberg (Eds.), *Lexical Cohesion and Corpus Linguistics* (pp. 65-83). Amsterdam, Netherlands: John Benjamins.

- Ciampi, F. (2005). Under Italian eyes: Conrad's critical reception in Italy. In M. Curreli (Ed.), *The Ugo Mursia Lectures: Papers from the International Conrad Conference, University of Pisa (September 16th-18th 2004)* (pp. 123-138). Pisa, Italy: Edizioni ETS.
- Clark, H. H., & Clark, E. V. (1977). Memory for prose. In H. H. Clark & E. V. Clark, *Psychology and Language: An Introduction to Psycholinguistics* (pp. 133-173). New York, NY: Harcourt Brace Jovanovich.
- Coats, J. M. (2013). Unreliable heterodiegesis and scientific racism in Conrad's *Secret Agent*. *Modernism/Modernity*, 20(4), 645-665.
- Conrad, J. (1950). Preface. In J. Conrad, *The Nigger of the "Narcissus" and Typhoon & Other Stories* (pp. vii-xii). London, UK: Dent & Sons LTD.
- Conrad, J. (1994). *Heart of Darkness*. London: Penguin. (Original work published 1902)
- Craig, H. (2008). 'Speak, that I may see thee': Shakespeare characters and common words. In P. Holland (Ed.), *Shakespeare Survey Online: Shakespeare, Sound and Screen*, 61 (pp. 281-288). Cambridge, UK: Cambridge University Press.
- Craig, H. (2015). Intelligent Archive [Computer software]. Newcastle, Australia: University of Newcastle.
- Culpeper, J. (2002a). A cognitive stylistic approach to characterisation. In E. Semino & J. Culpeper (Eds.), *Cognitive Stylistics: Language and Cognition in Text Analysis* (pp. 251-277). Amsterdam, Netherlands: John Benjamins.
- Culpeper, J. (2002b). Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*. In U. Melander-Marttala, C. Ostman, & M. Kyoto (Eds.), *Conversation in Life and Literature: Papers from the ASLA Symposium* (pp. 11-30). Uppsala, Sweden: Universitetsstryckeriet.
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*, 14(1), 29-59.
- Curle, R. (1914). *Joseph Conrad: A Study*. London, UK: Kegan Paul, Trench, Trübner.
- Curreli, M. (2009). *Le Traduzioni di Conrad in Italia*. Pisa, Italy: Edizioni ETS.
- Dancygier, B., & Sweetser, E. (2014). *Figurative Language*. Cambridge, UK: Cambridge University Press.

- Dixon, P., Bortolussi, M., Twilley, L. C., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics*, 22(1), 5-33.
- Dukāte, A. (2009). *Translation, Manipulation and Interpreting*. Frankfurt, Germany: Peter Lang.
- Egbert, J. (2012). Style in nineteenth century fiction: A multi-dimensional analysis. *Scientific Study of Literature*, 2(2), 167-198.
- Even-Zohar, I. (1990). Polysystem studies. *Poetics Today*, 11(1), 1-268.
- Firth, J. R. (1957). *Papers in Linguistics, 1934-1951*. London, UK: Oxford University Press.
- Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*. *International Journal of Corpus Linguistics*, 14(4), 492-523.
- Fischer-Starcke, B. (2010). *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. London, UK: Continuum.
- Fish, S. E. (1979). What is stylistics and why are they saying such terrible things about it? Part II. *Boundary 2*, 8(1), 129-146.
- Fish, S. E. (1980). What is stylistics and why are they saying such terrible things about it. In S. Fish, *Is There a Text in This Class? The Authority of Interpretative Communities* (pp. 68-96). Cambridge, MA: Harvard University Press.
- Flowerdew, J., & Mahlberg, M. (Eds.) (2009). *Lexical Cohesion and Corpus Linguistics*. Amsterdam, Netherlands: John Benjamins.
- Fowler, R. (1966). Linguistic theory and the study of literature. In R. Fowler (Ed.), *Essays on Style and Language: Linguistic and Critical Approaches to Literary Style* (pp. 1-28). London, UK: Routledge & K. Paul.
- Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press. 1996-2005. *Journal of English Linguistics*, 36(1), 5-38.
- GoGwilt, C. (2005). The interior: Benjaminian arcades, Conradian passages, and the 'impasse' of Jean Rhys. In P. Brooker & A. Thacker (Eds.), *Geographies of Modernism: Literatures, Cultures, Spaces* (pp. 65-75). London, UK: Routledge.
- Goonetilleke, D. C. R. A. (1999). Introduction. In D. C. R. A. Goonetilleke (Ed.), *Heart of Darkness* (pp. 9-48). Peterborough, UK: Broadview.

- Grabowski, Ł. (2013). Interfacing corpus linguistics and computational stylistics: Translation universals in translational literary Polish. *International Journal of Corpus Linguistics*, 18(2), 254-280.
- Granger, S., Lerot, J., & Petch-Tyson, S. (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam, Netherlands: Rodopi.
- Gries, S. T. (2010). Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily.... *International Journal of Corpus Linguistics*, 15(3), 327-343.
- Guerard, A. J. (1966). The journey within. In A. J. Guerard, *Conrad the Novelist* (pp. 1-59). Cambridge, MA: Harvard University Press.
- Gupta, K. (2013). *A corpus linguistic investigation into the media representation of the suffrage movement*. (Unpublished doctoral dissertation). University of Nottingham, Nottingham, UK.
- Halliday, M. A. K. (1971). Linguistic function and literary style: An enquiry into the language of William Golding's *The Inheritors*. In S. Chatman (Ed.), *Literary Style: A Symposium* (pp. 330-368). New York, NY: Oxford University Press.
- Halliday, M. A. K. (2004). *An Introduction to Functional Grammar - Third Edition*. London, UK: Arnold. (Original work published 1985)
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Longman.
- Hampson, R. (2005). Spatial stories: Joseph Conrad and James Joyce. In P. Brooker & A. Thacker (Eds.), *Geographies of Modernism: Literatures, Cultures, Spaces* (pp. 54-64). London, UK: Routledge.
- Hawkins, H. (2006). *Heart of Darkness* and racism. In P. B. Armstrong (Ed.), *Heart of Darkness* (pp. 365-375). New York, NY: Norton.
- Hegglund, J. (2005). Modernism, Africa and the myth of continents. In P. Brooker & A. Thacker (Eds.), *Geographies of Modernism: Literatures, Cultures, Spaces* (pp. 43-53). London, UK: Routledge.
- Hermans, T. (1999). *Translation in Systems. Descriptive and System-Oriented Approaches Explained*. Manchester, UK: St. Jerome.
- Ho, Y. (2011). *Corpus Stylistics in Principles and Practice: A Sylistic Exploration of John Fowles' The Magus*. London, UK: Continuum.

- Hoey, M. (1991). *Patterns of Lexis in Text*. London, UK: Oxford University Press.
- Holmes, J. S. (2004). The name and nature of translation studies. In L. Venuti (Ed.), *The Translation Studies Reader - Second Edition* (pp. 180-192). London, UK: Routledge. (Original work published 1972)
- Hoover, D. L. (2002). Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing*, 17(2), 157-180.
- Hoover, D. L. (2012). The Parallel Wordlist Spreadsheet. Retrieved from <https://files.nyu.edu/dh3/public/TheParallelWordlistSpreadsheet.html>
- Hoover, D. L. (2015). Cluster Analysis, Principal Component Analysis (PCA), and T-testing in Minitab. Retrieved from <https://files.nyu.edu/dh3/public/The%20Excel%20Text-Analysis%20Pages.html>
- Hori, M. (2004). *Investigating Dickens' Style. A Collocational Analysis*. Basingstoke, UK: Palgrave Macmillan.
- Huang, L. (2015). *Style in Translation: A Corpus-based Perspective*. Heidelberg, Germany: Springer.
- Huang, L., & Chu, C. (2014). Translator's style or translational style? A corpus-based study of style in translated Chinese novels. *Asia Pacific Translation and Intercultural Studies*, 1(2), 122-141.
- Hughes, G. (2006). *Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World*. Armonk, NY: M. E. Sharpe.
- Hung, J.-J., Bingenheimer, M., & Wiles, S. (2010). Quantitative evidence for a hypothesis regarding the attribution of early Buddhist translations. *Literary and Linguistic Computing*, 25(1), 119-134.
- Hunston, S. (2010). How can a corpus be used to explore patterns? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 152-166). New York, NY: Routledge.
- Minitab 16 Statistical Software [Computer software] (2014). State College, PA: Minitab, Inc.
- Jakobson, R. (1960). Closing statement: linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language* (pp. 350-377). Cambridge, MA: MIT Press.
- Jakubíček, M., Kilgarrieff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The

- TenTen corpus family. In A. Hardie & R. Love (Eds.), *Corpus Linguistics 2013 Abstract Book* (pp. 125-127). Lancaster, UK: UCREL, Lancaster University.
- Jarosz, L. (1992). Constructing the dark continent: Metaphor as geographic representation of Africa. *Geografiska Annaler. Series B, Human Geography*, 74(2), 105-115.
- Jeffries, L., & McIntyre, D. (2010). *Stylistics*. Cambridge, UK: Cambridge University Press.
- Johansson, S. (2003). Contrastive linguistics and corpora. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (pp. 31-44). Amsterdam, Netherlands: Rodopi.
- Kennedy, R. (2002). *Nigger: The Strange Career of a Troublesome Word*. New York, NY: Pantheon Books.
- Kenny, D. (2001). *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester, UK: St. Jerome.
- Kenny, D. (2008). Equivalence. In M. Baker & G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies* (pp. 96-99). New York, NY: Routledge.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1): 7-36.
- Kipling, R. (2013). The white man's burden. In T. Pinney (Ed.), *100 Poems Old and New: Rudyard Kipling* (pp. 111-113). Cambridge, UK: Cambridge University Press. (Original work published 1899)
- Knowles, O. (2009a). Critical responses: 1925-1950. In A. H. Simmons (Ed.), *Conrad in Context* (pp. 67-74). London, UK: Cambridge University Press.
- Knowles, O. (2009b). Literary influences. In A. H. Simmons (Ed.), *Joseph Conrad in Context* (pp. 33-41). London, UK: Cambridge University Press.
- Krein-Kühle, M. (2002). Cohesion and coherence in technical translation: The case of demonstrative reference. In L. V. Vaerenbergh (Ed.), *Linguistics and Translation Studies. Translation Studies and Linguistics. Linguistica Antverpiensia* (pp. 41-53). Antwerpen, Belgium: Hogeschool Antwerpen.
- Kruger, A., Wallmach, K., & Munday, J. (Eds.). (2013). *Corpus-based Translation Studies: Research and Applications*. London, UK: Bloomsbury.

- Kujawska-Lis, E. (2008). Turning *Heart of Darkness* into a racist text: A comparison of two Polish translations. *Conradiana*, 40(2), 165-178.
- Lambrou, M., & Stockwell, P. (Eds.) (2007). *Contemporary Stylistics*. London, UK: Continuum.
- Lavid, J., Arús, J., & Moratón, L. (2009). Comparison and translation: Towards a combined methodology for contrastive corpus studies. *International Journal of English Studies*, 9(3), 159-174.
- Laviosa, S. (2002). *Corpus-based Translation Studies*. Amsterdam, Netherlands: Rodopi.
- Laviosa, S. (2008). Corpus studies of translation universals: A critical appraisal. In A. Martelli & V. Pulcini (Eds.), *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat* (pp. 223-238). Monza, Italy: Polimetrica.
- Laviosa, S. (2013). Corpus-based translation studies: Where does it come from? Where is it going? In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based Translation Studies: Research and Applications* (pp. 13-32). London, UK: Bloomsbury.
- Lawtoo, N. (Ed.) (2012). *Conrad's Heart of Darkness and Contemporary Thought: Revisiting the Horror with Lacoue-Labarthe*. London, UK: Bloomsbury.
- Leavis, F. R. (1973). Joseph Conrad. In F. R. Leavis, *The Great Tradition. George Eliot, Henry James, Joseph Conrad* (pp. 173-226). London, UK: Chatto & Windus. (Original work published 1948)
- Leech, G. (1985). Stylistics. In T. A. van Dijk (Ed.), *Discourse and Literature. New Approaches to the Analysis of Literary Genres* (pp. 39-57). Amsterdam, Netherlands: John Benjamins.
- Leech, G., & Short, M. (2007). *Style in Fiction. A Linguistic Introduction to English Fictional Prose*. Harlow, UK: Pearson Education. (Original work published 1981)
- Lefevere, A. (1992). *Translation, Rewriting, and the Manipulation of Literary Fame*. London, UK: Routledge.
- Lefevere, A. (2014). Why waste our time on rewrites? The trouble with interpretation and the role of rewriting in an alternative paradigm. In T. Hermans (Ed.), *The Manipulation of Literature: Studies in Literary Translation* (pp. 215-243). New York: Routledge. (Original work published 1985)

- Levenson, M. (2009). Modernism. In A. H. Simmons (Ed.), *Joseph Conrad in Context* (pp. 179-186). London, UK: Cambridge University Press.
- Li, D., Zhang, C., & Liu, K. (2011). Translation style and ideology: A corpus-assisted analysis of two English translations of *Hongloumeng*. *Literary and Linguistic Computing*, 26(2), 153-166.
- Louw, W. E. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 157-174). Amsterdam, Netherlands: John Benjamins.
- MacDonald, M., & Hunter, D. (2013). The discourse of Olympic security: London 2012. *Discourse & Society*, 24(1), 66-88.
- Maczewski, J. M. (1996). Virginia Woolf's *The Waves* in French and German waters: A computer assisted study in literary translation. *Literary and Linguistic Computing*, 11(4), 175-186.
- Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach*. Amsterdam, Netherlands: John Benjamins.
- Mahlberg, M. (2007a). A corpus stylistic perspective on Dickens's *Great Expectations*. In M. Lambrou & P. Stockwell (Eds.), *Contemporary Stylistics* (pp. 19-31). London, UK: Continuum.
- Mahlberg, M. (2007b). Corpora and translations studies: Textual function of lexis in *Bleak House* and in a translation of the novel into German. In V. Intonti, G. Todisco, & M. Gatto (Eds.), *La Traduzione. Lo Stato dell'Arte. Translation. The State of the Art*. (pp. 115-135). Ravenna, Italy: Longo.
- Mahlberg, M. (2009). Lexical cohesion: Corpus linguistic theory and its application in English language teaching. In J. Flowerdew & M. Mahlberg (Eds.), *Lexical Cohesion and Corpus Linguistics* (pp. 103-122). Amsterdam, Netherlands: John Benjamins.
- Mahlberg, M. (2012). The corpus stylistic analysis of fiction or the fiction of corpus stylistics? In M. Huber & J. Mukherjee (Eds.), *Corpus Linguistics and Variation in English: Theory and Description* (pp. 77-95). Amsterdam, Netherlands: Rodopi.
- Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. New York, NY: Routledge.

- Mahlberg, M., Conklin, K., & Bisson, M.-J. (2014). Reading Dickens's characters: Employing psycholinguistic methods to investigate the cognitive reality of patterns in texts. *Language and Literature*, 23(4), 369-388.
- Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*, 4(2), 204-227.
- Mahlberg, M., & Smith, C. (2010). Corpus approaches to prose fiction: Civility and body language in *Pride and Prejudice*. In D. McIntyre & B. Busse (Eds.), *Language and Style* (pp. 449-467). Basingstoke, UK: Palgrave Macmillan.
- Mahlberg, M., & Smith, C. (2012). Dickens, the suspended quotation and the corpus. *Language and Literature*, 21(1), 51-65.
- McEney, T. (2009). Keywords and moral panics: Mary Whitehouse and media censorship. In D. Archer (Ed.), *What's in a Word-list? Investigating Word Frequency and Keyword Extraction* (pp. 93-124). Farnham, UK: Ashgate.
- McEney, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge, UK: Cambridge University Press.
- McIntyre, D. (2014). Characterisation. In P. Stockwell & S. Whiteley (Eds.), *The Cambridge Handbook of Stylistics* (pp. 149-164). Cambridge, UK: Cambridge University Press.
- McKenna, W., Burrows, J., & Antonia, A. (1999). Beckett's trilogy: Computational stylistics and the nature of translation. *Revue Informatique et Statistique dans les Sciences Humaines*, 35, 151-171. Retrieved from <http://promethee.philo.ulg.ac.be/RISSHpdf/Annee1999/Articles/WMcKennaet c.pdf>
- Miller, J. H. (2006). Should we read "Heart of Darkness"? In P. B. Armstrong (Ed.), *Heart of Darkness* (pp. 463-474). New York, NY: Norton.
- Miller, J. H. (2012). Prologue: Revisiting 'Heart of Darkness revisited' (in the company of Philippe Lacoue-Labarthe). In N. Lawtoo (Ed.), *Conrad's Heart of Darkness and Contemporary Thought: Revisiting the Horror with Lacoue-Labarthe* (pp. 17-35). London, UK: Bloomsbury.
- Mortara Gavarelli, B. (2010). *Manuale di Retorica*. Milano, Italy: Bompiani.
- Mouka, E., Saridakis, I. E., & Fotopoulou, A. (2015). Racism goes to the movies: A corpus-driven study of cross-linguistic racist discourse annotation and translation analysis. In C. Fantinuoli & F. Zanettin (Eds.), *New Directions in*

- Corpus-based Translation Studies* (pp. 35-69). Berlin, Germany: Language Science Press.
- Munday, J. (2008). *Style and Ideology in Translation. Latin America Writing in English*. London, UK: Routledge.
- Munday, J. (2011). Looming large: A cross-linguistic analysis of semantic prosodies in comparable reference corpora. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based Translation Studies: Research and Applications* (pp. 169-186). London, UK: Bloomsbury.
- Murfin, R. C. (Ed.) (1989). *Heart of Darkness: A Case Study in Contemporary Criticism*. Boston, MA: Bedford Books.
- Niland, R. (2009). Critical responses: 1950-1975. In A. H. Simmons (Ed.), *Joseph Conrad in Context* (pp. 75-82). Cambridge, UK: Cambridge University Press.
- O'Halloran, K. (2007). The subconscious in James Joyce's "Eveline": A corpus stylistic analysis that chews on the "Fish hook". *Language and Literature*, 16(3), 227-244.
- O'Prey, P. E. (1983). *Heart of Darkness*. London, UK: Penguin Books.
- Oakes, M. P., & Ji, M. (Eds.) (2012). *Quantitative Methods in Corpus-based Translation Studies*. Amsterdam, Netherlands: Benjamins.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London and New York: Routledge.
- Orpin, D. (2005). Corpus linguistics and critical discourse analysis. *International Journal of Corpus Linguistics*, 10(1), 37-61.
- Oster, U., & van Lawick, H. (2008). Semantic preference and semantic prosody: A corpus-based analysis of translation-relevant aspects of the meaning of phraseological units. In B. L.-T. Thelen (Ed.), *Translation and Meaning VIII* (pp. 333-344). Maastricht, Netherlands: Hogeschool Maastricht, School of Translation and Interpreting.
- Øverås, L. (1998). In search of the third code: An investigation of norms in literary translation. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 43(4), 557-570.
- Parks, T. (2007). *Translating Style: A Literary Approach to Translation, A Translation Approach to Literature*. Manchester, UK: St. Jerome.
- Partington, A. (2004). "Utterly content in each other's company": Semantic prosody

- and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131-156.
- Peters, J. G. (Ed.) (2008). *Conrad in the Public Eye*. Amsterdam, Netherlands: Rodopi.
- Purssell, A. (2009). Critical responses: 1975 - 2000. In A. H. Simmons (Ed.), *Joseph Conrad in Context* (pp. 83-90). Cambridge, UK: Cambridge University Press.
- Randall, J. H. (2008). Joseph Conrad: His outlook on life. In J. G. Peters (Ed.), *Conrad in the Public Eye* (pp. 108-123). Amsterdam, Netherlands: Rodopi. (Original work published 1925)
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549.
- Redelinghuys, K., & Kruger, H. (2015). Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics*, 20(3), 293-325.
- Rybicki, J. (2006). Burrowing into translation: Character idiolects in Henryk Sienkiewicz's Trilogy and its two English translations. *Literary and Linguistic Computing*, 21(1), 91-103.
- Rybicki, J., & Heydel, M. (2013). The stylistics and stylometry of collaborative translation: Woolf's *Night and Day* in Polish. *Literary and Linguistic Computing*, 28(4), 708-717.
- Rybicki, J., Hoover, D., & Kestemont, M. (2014). Collaborative authorship: Conrad, Ford and Rolling Delta. *Literary and Linguistic Computing*, 29(3), 422-431.
- Said, E. W. (1994). *Culture and Imperialism*. London, UK: Vintage.
- Saldanha, G. (2009). Principles of corpus linguistics and their application to translation studies research. *Revista Tradumàtica: Tecnologies de la Traducció*, 7. Retrieved from <http://webs2002.uab.es/tradumatica/revista/num2007/articles/2001/art.htm>
- Saldanha, G. (2011a). Translator style: Methodological considerations. *The Translator*, 17(1), 25-50.
- Saldanha, G. (2011b). Style of translation: The use of foreign words in translations by Margaret Jull Costa and Peter Bush. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based Translation Studies: Research and Applications* (pp. 237-258). London, UK: Bloomsbury.

- Saldanha, G. (2014). Style in, and of, translation. In S. Bermann & C. Porter (Eds.), *A Companion to Translation Studies* (pp. 95-106). West Sussex, UK: Wiley-Blackwell.
- Schäffner, C. (2008). Functionalist approaches. In M. Baker & G. Saldhana (Eds.), *Routledge Encyclopedia of Translation Studies* (pp. 115-121). New York, NY: Routledge.
- Scott, M. (2015a). *WordSmith Tools Help*. Liverpool, UK: Lexical Analysis Software. Retrieved from http://www.lexically.net/downloads/version6/HTML/index.html?getting_started.htm
- Scott, M. (2015b). *WordSmith Tools version 5* [Computer software]. Liverpool, UK: Lexical Analysis Software.
- Senn, W. (1980). *Conrad's Narrative Voice: Stylistic Aspects of his Fiction*. Bern, Switzerland: Francke Verlag.
- Simmons, A. H. (Ed.) (2009). *Joseph Conrad in Context*. Cambridge, UK: Cambridge University Press.
- Simpson, P. (2004). *Stylistics. A Resource Book for Students*. London, UK: Routledge.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press.
- Sinclair, J. (2004). *Trust the Text. Language, Corpus and Discourse*. London, UK: Routledge.
- Snell-Hornby, M. (1988). *Translation Studies: An Integrated Approach*. Amsterdam, Netherlands: John Benjamins.
- Spitzer, L. (1962). Linguistics and literary history. In L. Spitzer, *Linguistics and Literary History* (pp. 1-39). New York, NY: Russell & Russell Inc.
- Starcke, B. (2006). The phraseology of Jane Austen's *Persuasion*: Phraseological units as carriers of meaning. *ICAME Journal*, 30, 87-104.
- Stubbs, M. (2002). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford, UK: Blackwell.
- Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5-24.
- Thacker, A. (2005). The idea of critical literary geography. *New Formations*, 57, 56-

- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam, Netherlands: John Benjamins.
- Toolan, M. (2008). Narrative progression in the short story: First steps in a corpus stylistic approach. *Narrative*, 16(2), 105-120.
- Toury, G. (1995). *Descriptive Translation Studies – and Beyond*. Amsterdam, Netherlands: John Benjamins.
- Toury, G. (2004). The nature and role of norms in translation. In L. Venuti (Ed.), *The Translation Studies Reader (2nd. ed.)* (pp. 205-218). New York, NY: Routledge.
- Toury, G. (2014). A rationale for descriptive translation studies. In T. Hermans (Ed.), *The Manipulation of Literature: Studies in Literary Translation* (pp. 16-41). New York, NY: Routledge. (Original work published 1985)
- Trebits, A. (2009). Conjunctive cohesion in English language EU documents: A corpus-based analysis and its implications. *English for Specific Purposes*, 28(3), 199-210.
- Trupej, J. (2012). Translating racist discourse in Slovenia during the socialist period: Mark Twain's *Adventure of Huckleberry Finn*. In B. Fischer & M. N. Jensen (Eds.), *Translation and the Reconfiguration of Power Relations: Revisiting Role and Context of Translation and Interpreting* (pp. 91-107). Münster, Germany: Lit-Verlag.
- Turci, M. (2007). The meaning of 'dark*' in Joseph Conrad's *Heart of Darkness*. In D. R. Miller & M. Turci (Eds.), *Language and Verbal Art Revisited: Linguistic Approaches to the Study of Literature* (pp. 97-114). London, UK: Equinox.
- Tymoczko, M. (1998). Computerized corpora and the future of translation studies. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 43(4), 652-660.
- Tymoczko, M. (2010). *Translation, Resistance, Activism*. Amherst, MA: University of Massachusetts Press.
- Ulrych, M., & Anselmi, S. (2008). Towards a corpus-based distinction between language-specific and universal features of mediated discourse. In A. Martelli & V. Pulcini (Eds.), *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat* (pp. 257-273). Monza, Italy: Polimetrica.

- Van den Broeck, R. (2014). Second thoughts on translation criticism: A model of its analytic function. In T. Hermans (Ed.), *The Manipulation of Literature: Studies in Literary Translation* (pp. 54-62). New York, NY: Routledge. (Original work published 1985)
- Venuti, L. (2008). *The Translator's Invisibility: A History of Translation* (2nd ed.). London, UK: Routledge.
- Venuti, L. (Ed.) (2004). *The Translation Studies Reader* (2nd ed.). London, UK: Routledge.
- Vermeule, B. (2010). *Why Do We Care about Literary Characters?* Baltimore, MD: The Johns Hopkins University Press.
- Wales, K. (2001). *A Dictionary of Stylistics*. Harlow, UK: Pearson Education.
- Wang, Q., & Li, D. (2012). Looking for translator's fingerprints: A corpus-based study on Chinese translation of *Ulysses*. *Literary and Linguistic Computing*, 27(1), 81-93.
- Warren, M. (2009). Cohesive chains and speakers' choice of prominence. In J. Flowerdew & M. Mahlberg (Eds.), *Lexical Cohesion and Corpus Linguistics* (pp. 45-63). Amsterdam, Netherlands: John Benjamins.
- Watt, I. (2006). Impressionism and symbolism in *Heart of Darkness*. In P. B. Armstrong (Ed.), *Heart of Darkness* (pp. 349-365). New York, NY: Norton.
- Watts, C. E. (2008). *Heart of Darkness and Other Tales*. Oxford, UK: Oxford University Press. (Original work published 1990)
- Waugh, A. (1919). *Tradition and Change: Studies in Contemporary Literature*. London, UK: Chapman and Hall.
- Winters, M. (2005). *A Corpus-based Study of Translator Style: Oeser's and Orth Guttman's German Translations of F. Scott Fitzgerald's The Beautiful and Damned*. Dublin City University, Dublin, Ireland.
- Winters, M. (2007). F. Scott Fitzgerald's *Die Schönen und Verdamnten*: A corpus-based study of speech-act report verbs as a feature of translators. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 52(3), 412-425.
- Winters, M. (2010). From modal particles to point of view: A theoretical framework for the analysis of translator attitude. *Translation and Interpreting Studies*, 5(2), 163-185.
- Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy:

A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129.

Youmans, G. (1990). Measuring lexical style and competence: The type-token vocabulary curve. *Style*, 24(4), 584-599.

Zins, H. (1982). *Joseph Conrad and Africa*. Nairobi, Kenya: Kenya Literature Bureau.

APPENDICES

Appendix 1. Texts in the reference corpora

1. Reference Corpus (RC)

- 1887 – *Little Novels*. Collins, W.
- 1889 – *The Legacy of Cain*. Collins, W.
- 1889 – *The Master of Ballantrae*. Stevenson, R.L.
- 1893 – *Catriona*. Stevenson, R.L.
- 1894 – *Life's Little Ironies*. Hardy, T.
- 1894 – *The Jungle Book*. Kipling, R.
- 1894 – *The People of the Mist*. Haggard, H.R.
- 1895 – *Jude the Obscure*. Hardy, T.
- 1895 – *The Amazing Marriage*. Meredith, G.
- 1896 – *The Island of Doctor Moreau*. Wells, H.G.
- 1896 – *The London Pride*. Braddon, M.E.
- 1898 – *The War of the Worlds*. Wells, H.G.
- 1901 – *Kim*. Kipling, R.
- 1902 – *The Hound of the Baskervilles*. Conan Doyle, A.
- 1904 – *Henry Brocken*. De la Mare, W.
- 1905 – *Ayesha, The Return of She*. Haggard, H.R.
- 1905 – *The Marriage of William Ashe*. Ward, H.
- 1905 – *Where Angels Fear to Tread*. Forster, E.M.
- 1906 – *The Fifth Queen*. Madox Ford, F.
- 1906 – *The Man of Property*. Galsworthy, J.
- 1908 – *A Room with a View*. Forster, E.M.
- 1908 – *The Old Wives' Tales*. Bennett, A.
- 1910 – *Celt and Saxon*. Meredith, G.
- 1910 – *Clayhanger*. Bennett, A.
- 1910 – *The Return*. De la Mare, W.
- 1911 – *In a German Pension*. Mansfield, K.
- 1911 – *The Case of Richard Meynell*. Ward, H.
- 1913 – *Sons and Lovers*. Lawrance, D.H.
- 1914 – *Innocent, Her Fancy and His Fact*. – Corelli, M.
- 1915 – *The Good Soldier*. Madox Ford, F.

1915 – *The Valley of Fear*. Conan Doyle, A.
1918 – *The Clue of the Twisted Candle*. Wallace, E.
1919 – *Night and Day*. Woolf, V.
1920 – *Women in Love*. Lawrence, D.H.
1920 – *The Mysterious Affair at Style*. Christie, A.
1921 – *Bones in London*. Wallace, E.
1921 – *Crome Yellow*. Huxley, A.
1921 – *To Let*. Galsworthy, J.
1921 – *The Secret Power*. Corelli, M.
1922 – *Jacob's Room*. Woolf, V.
1922 – *The Garden Party and Other Stories*. Mansfield, K.
1922 – *The Secret Adversary*. Christie, A.

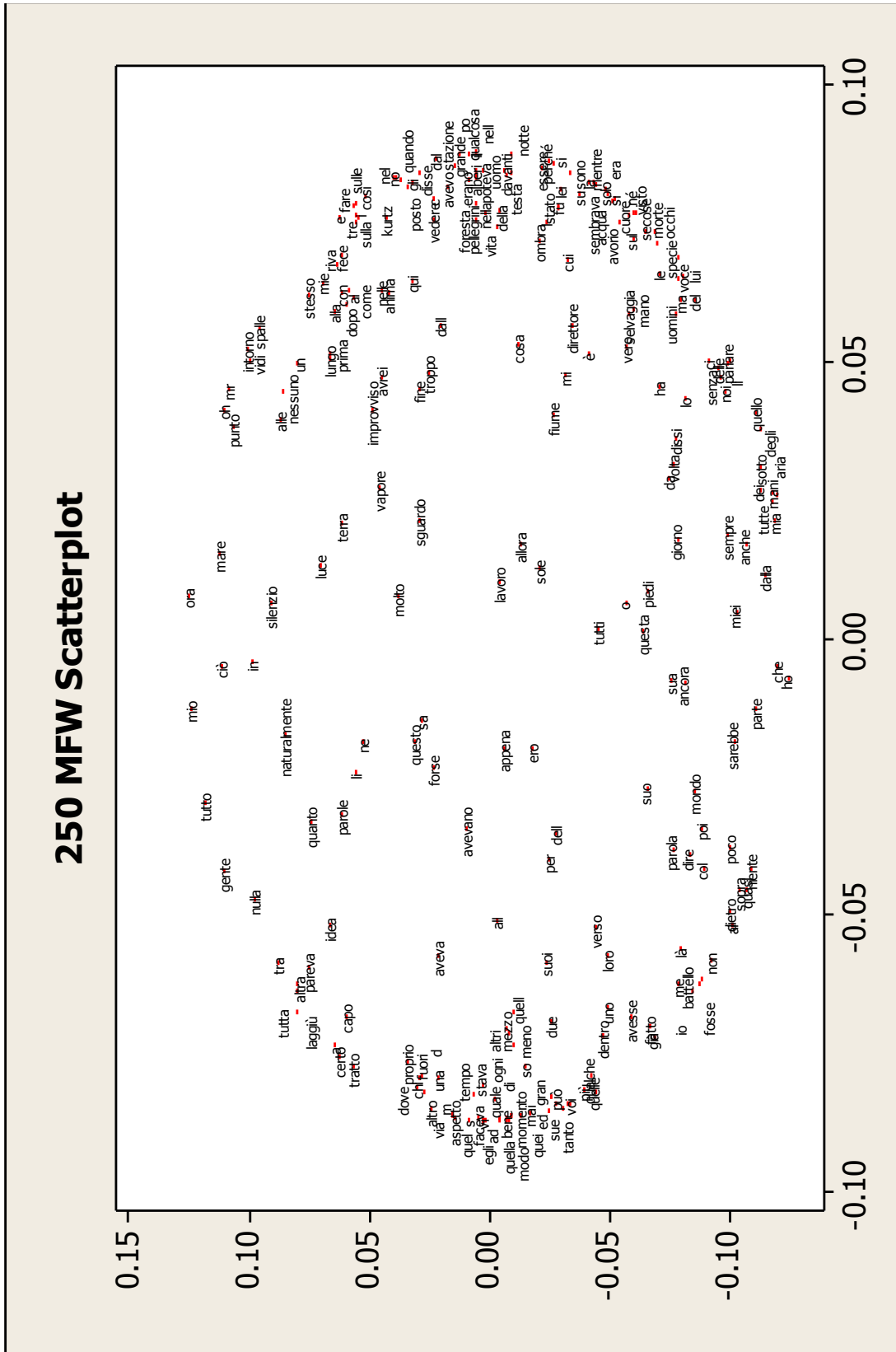
2. Conrad Corpus (CC)

1985 – *Almayer's Folly*
1896 – *An Outcast of the Islands*
1897 – *The Nigger of the "Narcissus"*
1898 – *Tales of Unrest*
1898 – *Youth*
1900 – *Lord Jim*
1901 – *Amy Foster*
1901 – *Falk*
1902 – *The End of the Tether*
1902 – *To-morrow*
1902 – *Typhoon*
1904 – *Nostramo*
1907 – *The Secret Agent*
1908 – *A Set of Six*
1911 – *Under Western Eyes*
1912 – *'Twixt Land and Sea*
1913 – *Change*
1915 – *Victory*
1915 – *Within the Tides*
1917 – *The Shadow Line*
1919 – *The Arrow of Gold*

1920 – *The Rescue*

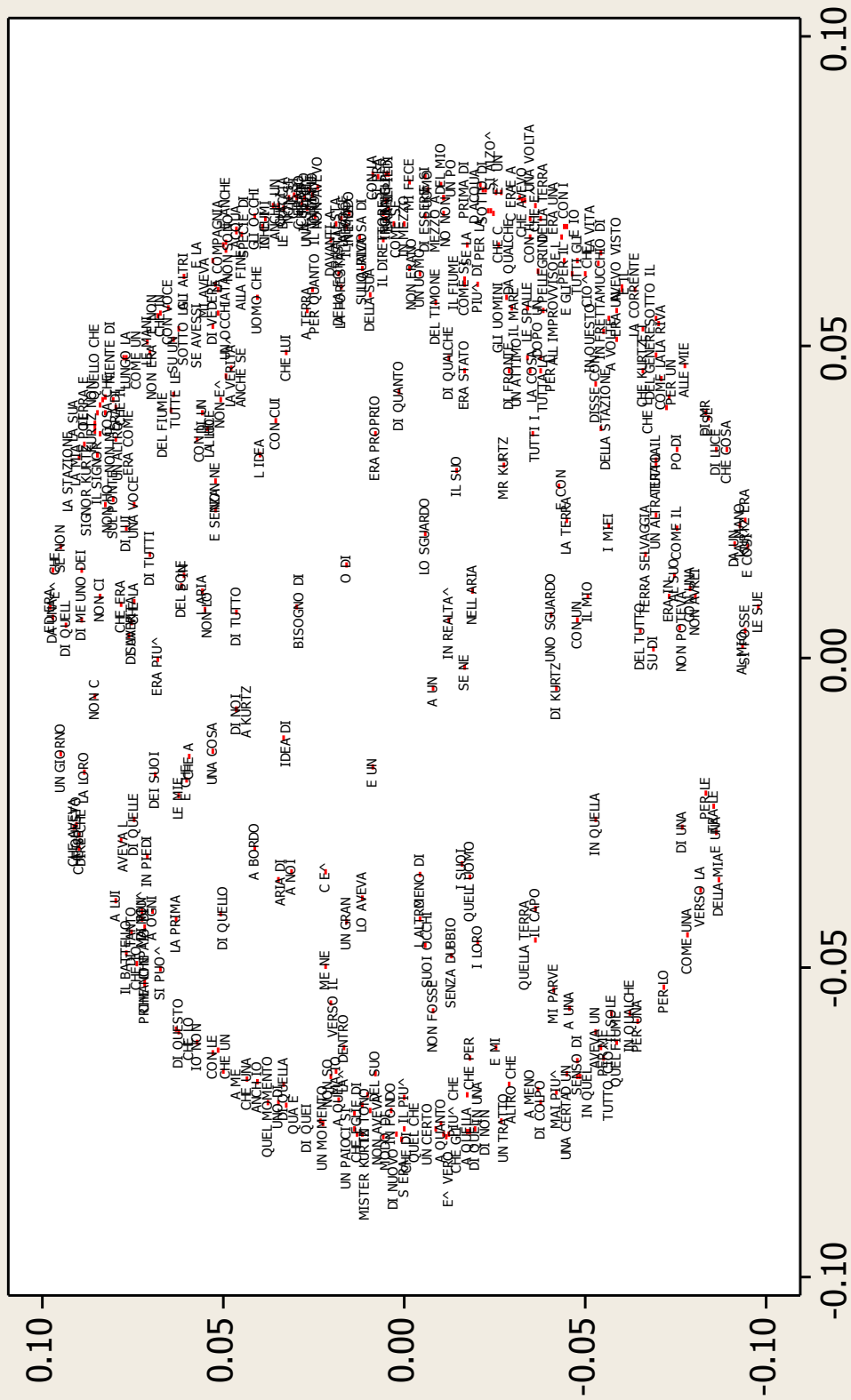
1925 – *Tales of Hearsay*

Appendix 2. Larger scatterplots



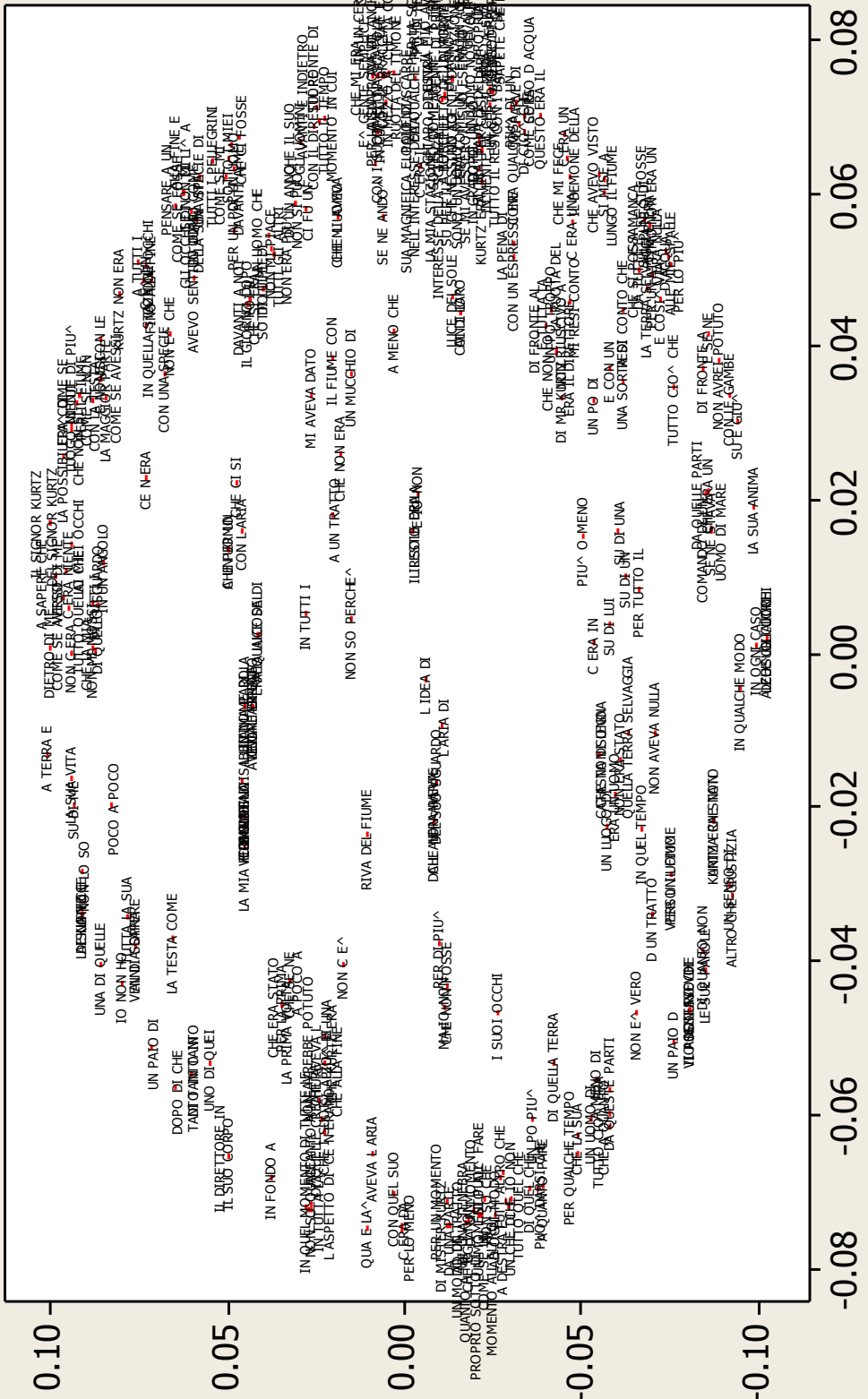
1. Scatterplot based on the 250 MFWs of the TTs

350 Two-word MFS Scatterplot



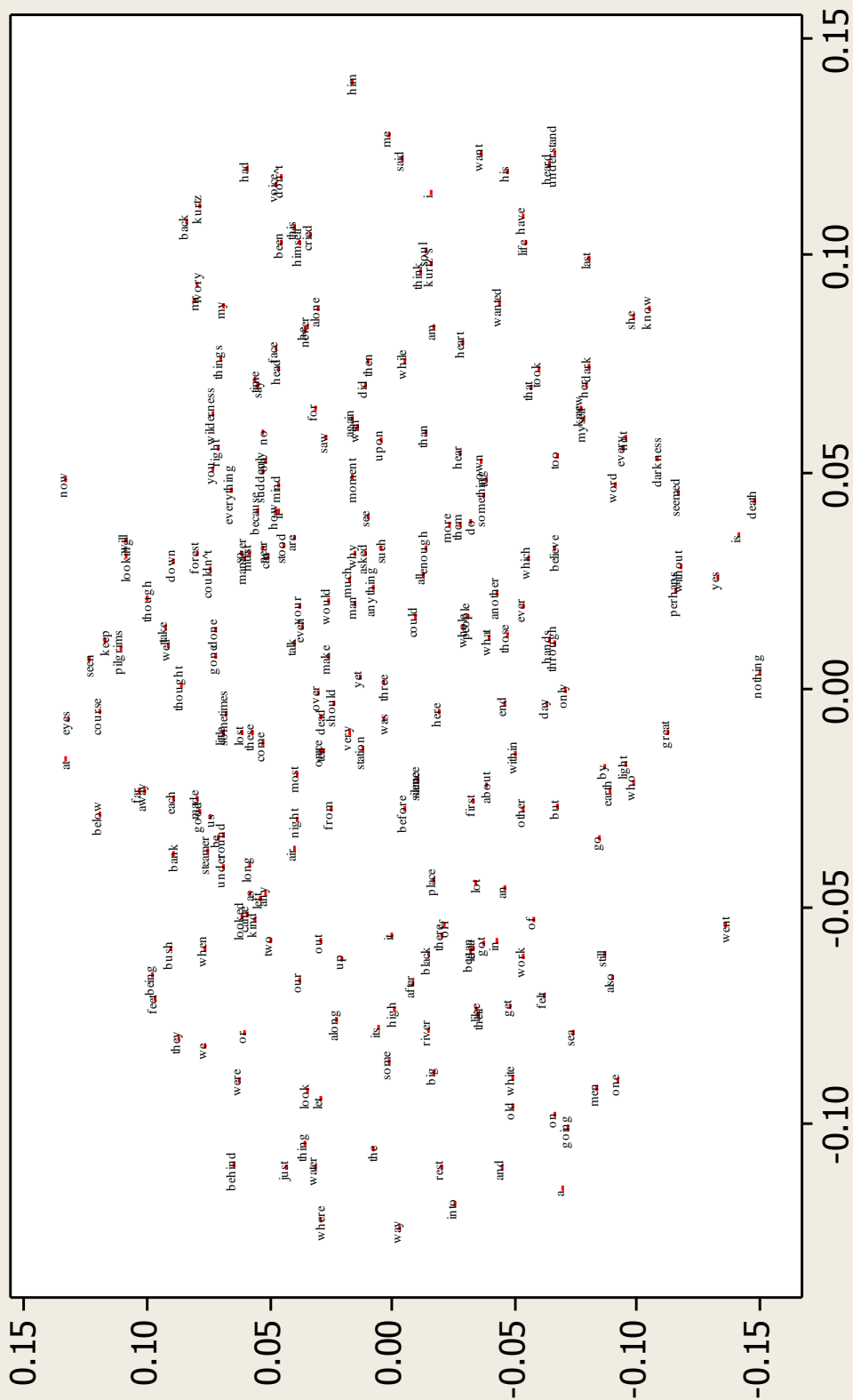
2. Scatterplot based on the 350 two-word MFSs of the TTs

350 Three-word MFS Scatterplot



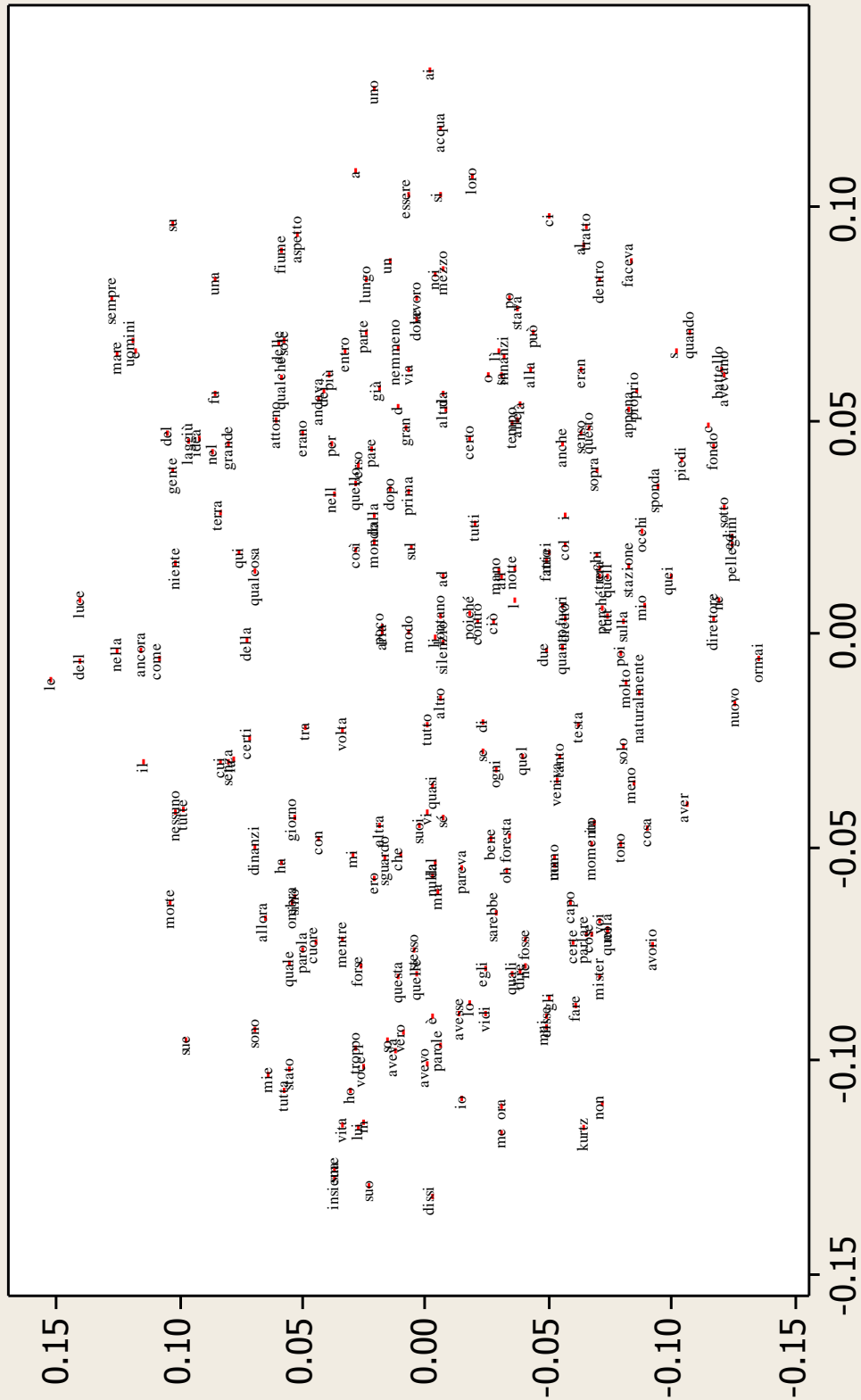
3. Scatterplot based on the 350 three-word MFSs of the TTs

250 MFW HoD Sections Scatterplot



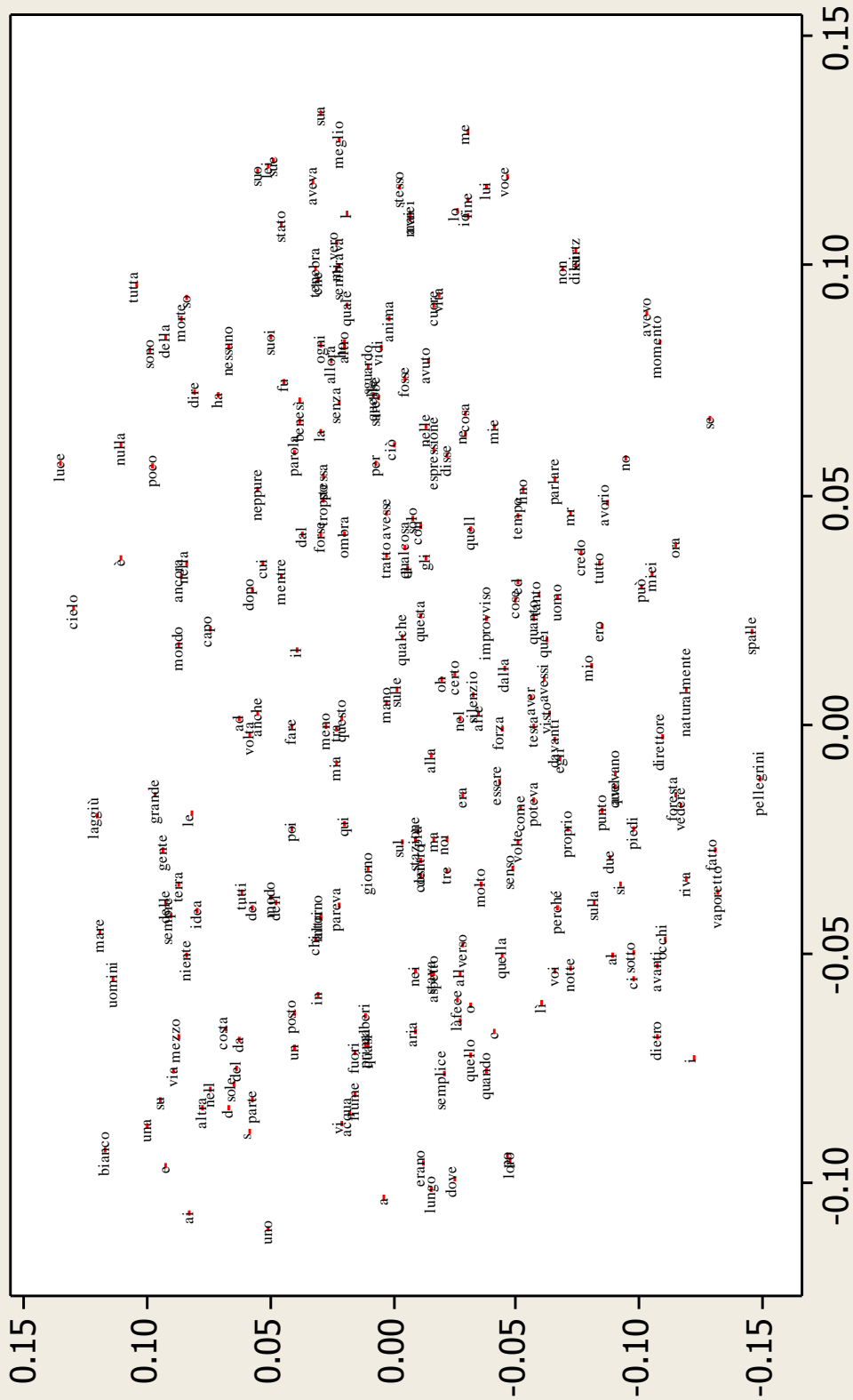
4. HoD sections scatterplot based on 250 MFWs

250 MFW Translation S Sections Scatterplot



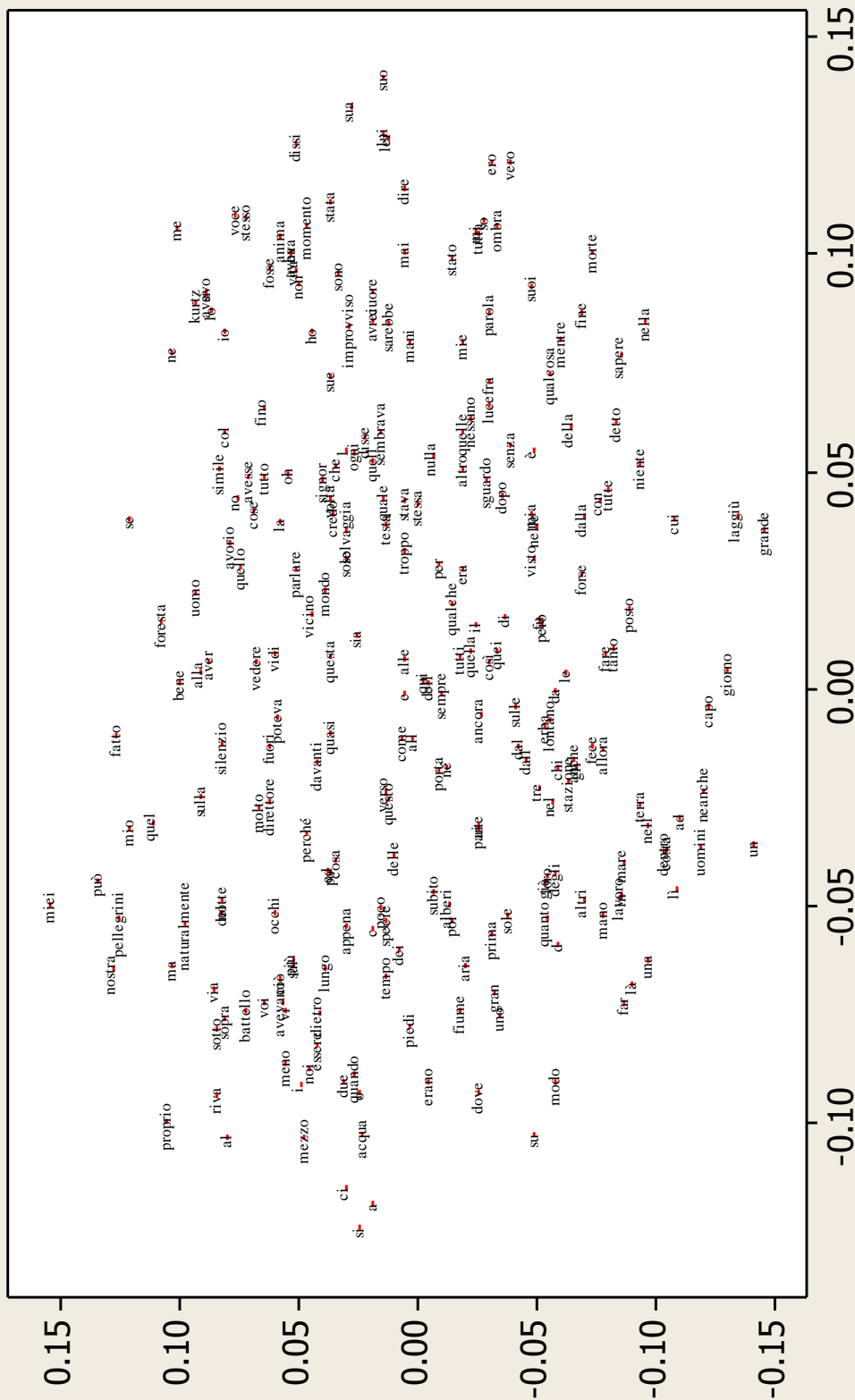
5. Translation S sections scatterplot based on 250 MFWs

250 MFW Translation B Sections Scatterplot

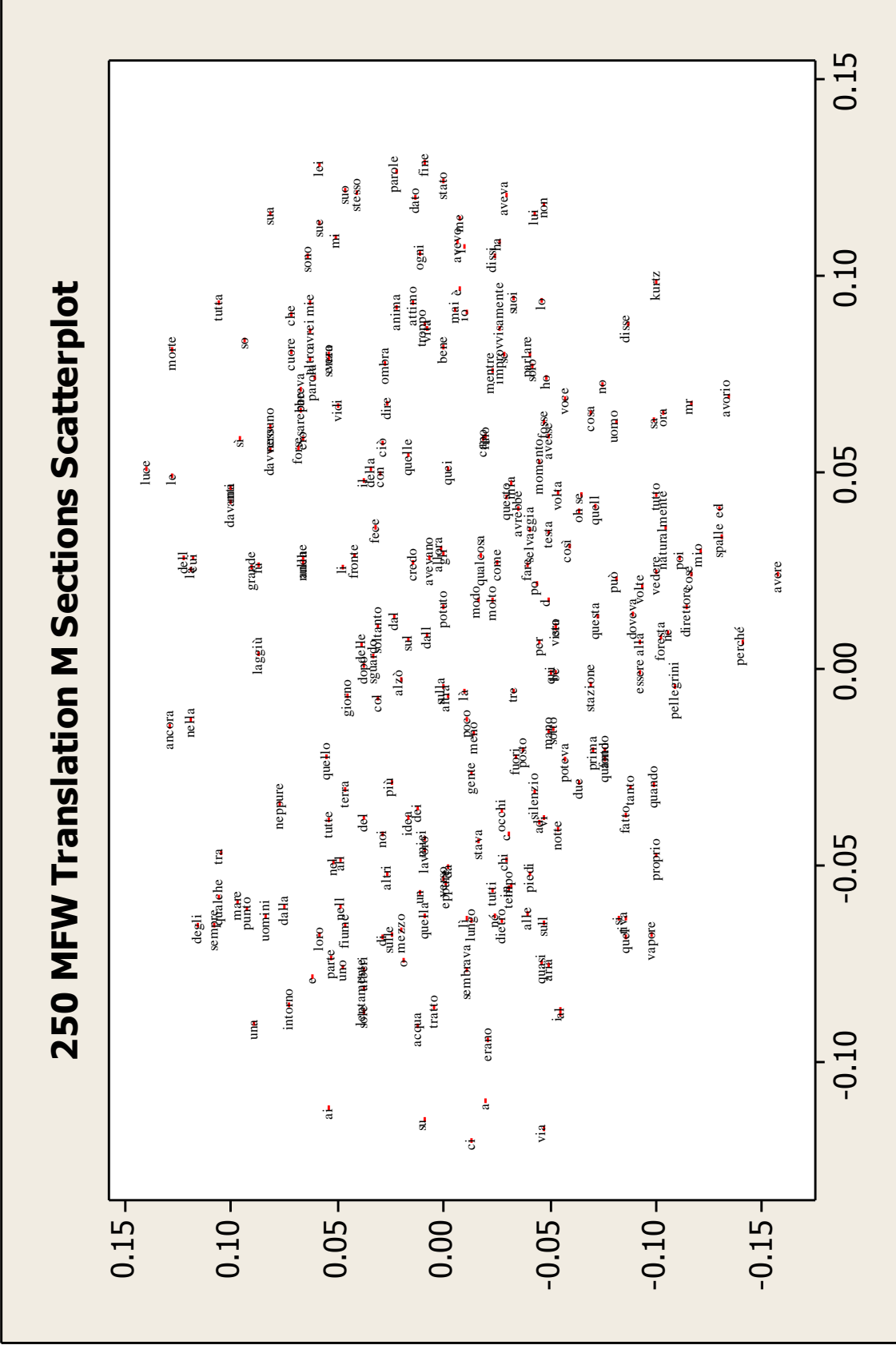


6. Translation B sections scatterplot based on 250 MFWs

250 MFW Translation G Sections Scatterplot



7. Translation G sections scatterplot based on 250 MFWs



8. Translation M sections scatterplot based on 250 MFWs

Appendix 3. Sections comparison

Section	Tokens	Types	TTR	STTR	Sentences	Mean (in words)
HoD Section (1)	3876	1273	32.83	44.57	198.00	19.58
S Section (1)	4161	1618	38.88	52.50	196.00	21.23
B Section (1)	3861	1558	40.34	52.53	202.00	19.12
G Section (1)	3889	1525	39.20	51.13	207.00	18.79
M Section (1)	3723	1475	39.61	51.60	200.00	18.62
HoD Section (2)	3894	1322	33.94	47.50	246.00	15.83
S Section (2)	4252	1729	40.65	55.60	230.00	18.49
B Section (2)	3797	1605	42.26	55.23	249.00	15.25
G Section (2)	3857	1620	41.99	54.67	248.00	15.56
M Section (2)	3717	1581	42.52	55.83	245.00	15.18
HoD Section (3)	3880	1199	30.90	46.00	281.00	13.81
S Section (3)	4209	1593	37.84	54.30	251.00	16.77
B Section (3)	3731	1457	39.04	54.10	283.00	13.19
G Section (3)	3812	1493	39.16	54.07	269.00	14.17
M Section (3)	3650	1413	38.70	53.27	269.00	13.57
HoD Section (4)	3863	1192	30.86	44.20	262.00	14.74
S Section (4)	4114	1570	38.16	53.35	259.00	15.88
B Section (4)	3738	1453	38.87	53.27	267.00	14.00
G Section (4)	3866	1465	37.89	52.03	255.00	15.16
M Section (4)	3734	1421	38.06	52.67	258.00	14.47
HoD Section (5)	3870	1282	33.12	47.23	218.00	17.76
S Section (5)	4351	1674	38.47	53.90	221.00	19.69
B Section (5)	3759	1549	41.20	54.70	222.00	16.94
G Section (5)	3898	1529	39.22	53.53	224.00	17.41
M Section (5)	3846	1548	40.25	54.23	219.00	17.56
HoD Section (6)	3889	1236	31.78	45.37	214.00	18.17
S Section (6)	4154	1591	38.30	52.83	210.00	19.78
B Section (6)	3805	1503	39.50	52.77	217.00	17.53
G Section (6)	3911	1497	38.28	51.67	219.00	17.86
M Section (6)	3818	1527	39.99	53.63	212.00	18.01
HoD Section (7)	3876	1241	32.01	45.13	291.00	13.32
S Section (7)	4265	1633	38.28	53.10	273.00	15.63
B Section (7)	3810	1514	39.73	53.97	294.00	12.96
G Section (7)	3842	1510	39.29	53.40	287.00	13.39
M Section (7)	3734	1510	40.43	54.17	284.00	13.15
HoD Section (8)	3861	1148	29.73	43.23	290.00	13.32
S Section (8)	4099	1520	37.07	52.53	268.00	15.30
B Section (8)	3673	1433	39.00	52.67	294.00	12.50
G Section (8)	3772	1416	37.53	51.07	280.00	13.47
M Section (8)	3586	1385	38.61	51.63	280.00	12.81
HoD Section (9)	3867	1268	32.78	45.70	253.00	15.29
S Section (9)	4057	1611	39.71	55.33	241.00	16.83
B Section (9)	3723	1547	41.54	55.97	258.00	14.43
G Section (9)	3811	1523	39.95	53.67	250.00	15.25

M Section (9)	3576	1456	40.70	54.87	248.00	14.42
HoD Section (10)	3873	1140	29.43	43.70	272.00	14.24
S Section (10)	4132	1430	34.61	49.72	268.00	15.42
B Section (10)	3811	1393	36.55	51.37	274.00	13.91
G Section (10)	3828	1332	34.80	49.33	276.00	13.87
M Section (10)	3636	1315	36.17	50.53	273.00	13.32