New Frontiers in Translation Studies

Sara Laviosa
Adriana Pagano
Hannu Kemppanen
Meng Ji

Textual and Contextual Analysis in Empirical Translation Studies

Springer

New Frontiers in Translation Studies

Series editor

Defeng Li
Centre for Translation Studies, SOAS, University of London,
London, UK

Centre for Studies of Translation, Interpreting and Cognition,
University of Macau, Macau SAR

More information about this series at http://www.springer.com/series/11894

Sara Laviosa · Adriana Pagano
Hannu Kemppanen · Meng Ji

Textual and Contextual Analysis in Empirical Translation Studies

Springer

Sara Laviosa
Dipartimento LELIA
University of Bari Aldo Moro
Bari
Italy

Adriana Pagano
Translation Studies
Federal University of Minas Gerais
Belo Horizonte
Brazil

Hannu Kemppanen
Translation Studies
University of Eastern Finland
Joensuu
Finland

Meng Ji
The University of Sydney
Sydney, NSW
Australia

ISSN 2197-8689 ISSN 2197-8697 (electronic)
New Frontiers in Translation Studies
ISBN 978-981-10-1967-8 ISBN 978-981-10-1969-2 (eBook)
DOI 10.1007/978-981-10-1969-2

Library of Congress Control Number: 2016944846

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Contents

Introduction

Textual and Contextual Analysis in Empirical Translation Studies is a volume co-authored by world's leading scholars of empirical translation studies. Since the early 1990s, corpus translation research has acted as a stimulus to empirical studies of translation. With the rapid development of large-scale databases and computational tools for natural language processing, corpus translation research has become mainstream in theoretical- and practical-oriented translation studies. As with many new disciplines characterized by the use of digital resources and technologies, there have been heated debates over the distinction between specific sets of research methodologies that have come to diversify corpus translation research, such as corpus-based, corpus-assisted, corpus-oriented or corpus-driven. A central aim of this book was therefore to explore the viability and productivity of integrating distinct yet interrelated research methods in recent empirical translation studies. The papers in this book demonstrate how the collection, quantitative processing and qualitative analysis of corpus data yield valuable insights into the dynamics of translation activities and products in particular social and cultural backgrounds.

From the translation of well-established English loanwords in Italy and the translation of foreign literature in Brazilian's Golden Age in the 1940s to the translation of political history in Finland during the Soviet Era and, more recently, the translation and adaptation of English environmental news on climate change in mainland China, this book covers latest empirical findings of the translation studies in Europe, Latin America and the Asia-Pacific. Our joint efforts are geared towards bridging the long-existing gap in corpus translation research, i.e. the gap between corpus-based textual analysis and the analysis of the social and cultural context which gives rise to particular intellectual products such as translations. The case studies show the divide between textual and contextual analyses in current corpus research can be closed by exploring the potential of existing and novel quantitative methods adapted from related fields of enquiry.

The central argument made in this book by four world's leading translation scholars is that current translation studies especially the corpus-orientated branch represent a mixed use of qualitative and quantitative analysis (Chaps. 1–3), and an important feature of latest research developments is the exploration of the relationship between detected translational features and contextual variables to fill a critical gap in current quantitative translation studies. To be specific, the opening chapter by Laviosa gives an overview of the growth of empirical translation studies since the 1990s. This is followed by detailed review and discussion in Chaps. 2 and 3 by Kemppanen of important analytical concepts developed in early empirical translation studies such as keyword analysis. Chapter 4 (Pagano) and Chap. 5 (Ji) focus on the introduction and use of exploratory statistics in quantitative translation studies. These include descriptive analysis such as keyword list generation and multivariate analyses (MVA) such as cluster analysis, principal component analysis and confirmatory MVA. There is no doubt, in our view, that the establishment of these new analytical techniques for the empirical investigation of translation will receive some criticism and become the object of scholarly debate, as was the case with the development of corpus translation studies since the 1990s and, indeed, the growth of translation studies as a broad interdisciplinary and increasingly empirical field of study in its own right.

Chapter 1
Empirical Translation Studies: From Theory to Practice and Back Again

1.1 Introduction[1]

When corpora began to be used in a systematic way for the empirical study of translation, Tymoczko (1998: 657) claimed that the appeal of corpus studies lay in their potential "to illuminate both similarity and difference and to investigate in a manageable form the particulars of language-specific phenomena of many different languages and cultures". Today, the envisioned role of corpora as invaluable repositories of data for carrying out contrastive analyses across languages and cultures is a reality in descriptive as in applied studies. In this chapter I first give an overview of the evolution of corpus studies of translation from their introduction in the discipline to current research endeavours. Next, I examine the holistic approach to translating cultural difference put forward by Tymoczko (2007). I also discuss the role that corpora can play in raising awareness of "the largest elements of cultural difference that separate the source culture and the target culture as a framework for coordinating the particular decisions about culture that occur as the text is actually transposed into the target language" (ibid: 235). Finally, I consider the application

[1]This chapter draws largely on five keynote papers presented by the author from 2008 to 2014. "Empirical Translation Studies: From Theory to Practice and Back Again". *New Perspectives in Translation Studies*, Ningbo University, China, 13–16 June 2014. "Corpora and Holistic Cultural Translation" *I Coloquio Hermēneus. Los estudios de Traducción e Interpretación basados en corpus*, Facultad de Traducción e Interpretación de Soria (Universidad de Valladolid), 26–27 March 2014. http://wp.me/P3SeK5-4I. "A transcultural Conceptual Framework for Corpus-based Translation Pedagogy" *Using Corpora in Contrastive and Translation Studies 2010 Conference* (*UCCTS2010*), Edge Hill University, Ormskirk, UK, 27–29 July. http://www.lancs.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/. "Corpus-based Translation studies: Theory, Findings, Applications". Guest lecture given at the Department of Professional and Intercultural Communication, Norwegian School of Economics and Business Administration, Bergen, Norway, 9 October 2009. "Discovery and Justification Procedures in the Corpus-based Translation Classroom". *Translation Challenges: From Training to Profession*, Hammamet, Tunisia, 28–29 November 2008. http://translationinfo.webs.com/abstracts.htm.

S. Laviosa et al., *Textual and Contextual Analysis in Empirical Translation Studies*, New Frontiers in Translation Studies,
DOI 10.1007/978-981-10-1969-2_1

of Tymoczko's approach in translator education with a view to underscoring the interplay between theory, description and practice within the discipline as a whole.

1.2 Introducing Corpora in Translation Studies: 1993–2003

The introduction of corpora in Translation Studies was conceived within an empirical paradigm and came to be as a result of the convergence between the discovery and justification procedures put forward by Gideon Toury (1995/2012) for the study of translation and the data-driven approach developed by Corpus Linguistics for the study of languages. The synergy between Descriptive Translation Studies and Corpus Linguistics acted as a stimulus to the creation of a variety of corpus resources, the development of a descriptive research methodology and the growth of a line of enquiry that was put forward in the 1980s and gathered momentum thanks to the availability of corpora. This body of research is known as the quest for translation universals, which are posited as probabilistic laws of translational behaviour (Toury 1995/2012). I will now give some details about each of these achievements of CTS.

One of the first corpus resources designed for contrastive linguistics and translation studies is the *English-Norwegian Parallel Corpus* (ENPC). It was compiled at the University of Oslo under the direction of Stig Johansson and served as a model for the bidirectional parallel corpus of English and Portuguese, COMPARA.[2] Another corpus design is the monolingual comparable corpus. An example is the *Translational English Corpus* (TEC).[3] It was created at the University of Manchester under the direction of Mona Baker. Another example is the *Corpus of Translated Finnish* (CTF). It was compiled at the Savonlinna School of Translation Studies by Anna Mauranen's research group. The methodology adopted by CTS involves a helical progression from the elaboration of descriptive, interpretive and explanatory hypotheses to inferences about the non-observable culturally-determined norms that govern translators' choices.

Initially, research focused on four universals: simplification, explicitation, the law of growing standardization (largely compatible with normalization) and the law of interference. Simplification is "the process and/or result of making do with less words" (Blum-Kulka and Levenston 1983: 119). Explicitation is "an observed cohesive explicitness from SL to TL texts regardless of the increase traceable to differences between the linguistic and textual systems involved" (Blum-Kulka 1986: 19). The law of growing standardization posits that "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target

[2]Available at: http://www.linguateca.pt/COMPARA/Welcome.

[3]Available at: http://www.llc.manchester.ac.uk/ctis/research/english-corpus/.

repertoire" (Toury 1995/2012: 304). The law of interference states that "in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text" (Toury 1995/2012: 310). In sum, during the first decade of its life, CTS built upon, refined, extended and diversified previous research into the regularities of translational language.

Meanwhile, corpora were making inroads into Applied Translation Studies. In this area of research and practice, corpora were used mainly as translation aids in translator training. Corpora were utilized as repositories of data for retrieving translation equivalents, acquiring content knowledge about specialized subject fields and developing stylistic fluency and terminological accuracy in the target language. Translation pedagogy drew mainly on Data-Driven Learning (DDL), developed by Tim Johns for the teaching of languages (Johns 1991a, b), and on constructivist principles, which constitute a dominant paradigm in contemporary educational philosophy and "serve as a strong cornerstone for the development of student- and praxis-relevant teaching methods" (Kiraly 2003: 8).

More specifically, the DDL approach adopts the principles of Corpus Linguistics and involves carrying out small-scale projects where students identify problem areas arising from translation practice, suggest hypotheses and then test them with their own tutor who has the role of "director and coordinator of student-initiated research" (Johns 1991a: 3). The approach adopted by the collaborative-constructivist method combines social constructivism with modern functionalist theories and expertise studies. The design involves collaborative learning and project-based activities. The procedure requires that students engage in an authentic or realistically simulated translation project together with peers (Kiraly 2000, 2003). Summing up, within the empirical paradigm, which can be regarded, in line with Chesterman (1998), as the most important trend that characterized Translation Studies in the 1990s, corpora engendered a number of novel syntheses in the pure and applied branches of the discipline.

1.3 Consolidating Corpora in Translation Studies: 2003–2013

The second decade in the life of CTS is marked by two international conferences entirely devoted to corpora and Translation Studies. The first was held in Pretoria in 2003, it was entitled *Corpus-based Translation Studies: Research and Applications* (Kruger et al. 2011). The second was hosted in Shanghai in 2007, *Conference and Workshop on Corpora and Translation Studies*. Of note is also the establishment of a strong partnership between contrastive and translation studies, in keeping with the research programme initiated by Stig Johansson in the 1990s and pursued in several interdisciplinary collected volumes such as Granger et al. (2003). The cooperation between these two disciplines finds its voice in a series of biennial international conferences, *Using Corpora in Contrastive and Translation Studies* (UCCTS). The first conference was held at Zhejiang University in Hangzhou, China, on 25–27 September 2008 (Xiao 2010). The second one, jointly organized by Edge Hill University, the University of Bologna and Beijing Foreign Studies University, took place at Edge Hill University in Ormskirk, UK, on 27–29 July 2010. The third UCCTS conference was held at Lancaster University, UK, from 24 to 26 July 2014.

As regards Descriptive Translation Studies, many new corpora have been created in the last ten years or so, as amply testified by Federico Zanettin's web page[4] and his monograph (Zanettin 2012). A novelty is the design of corpora of interpreted speeches, the first being the European Parliament Interpreting Corpus (EPIC),[5] and the consequent growth of a new body of research named Corpus-based Interpreting Studies (CIS) (see Setton 2011 for a review). Its main goal is to unearth the specificity of interpreting vis-à-vis original oral discourse and written translation in the same target language. This is a line of enquiry that was first proposed and pursued by Shlesinger (1998, 2009). The creation of new corpus resources goes hand in hand with the development of methodology and statistical analyses. These are enriched by contextual data and are becoming more and more sophisticated thanks to the advancement of technology. As for the range of research endeavours, the quest for translation universals goes on (Mauranen and Kujamäki 2004; Russo et al. 2006; Espunya 2007; Bruti and Pavesi 2008; Steiner 2005, 2008; Ulrych and Anselmi 2008; Gaspari and Bernardini 2010; Xiao et al. 2010; Zuffery and Cartoni 2014). In addition to the four universals mentioned earlier, a new one emerged, namely the Unique Items Hypothesis (UIH). It states that target-language-specific elements, which do not have equivalents in the source language, tend to be under-represented in translated texts, since "they do not readily suggest themselves as translation equivalents" (Tirkkonen-Condit 2004: 177–178).

Despite growing interest in this ambit of research, a number of scholars expressed serious criticism about the tenability of the very concept of translation universal. House (2008), for example, claimed that the investigation of universals is futile since there are no and there can be no translation-inherent universals. The reasons for denying the existence of linguistic features of translation per se are as follows: (a) since translation is an act that operates on language, the universals of language also apply to translation; (b) translation is inherently language-pair specific, hence even corpus-based multi-pair comparisons remain agglomerations of different pairs; (c) the suggested candidates for the status of translation universal for one particular translation direction need not necessarily be candidates for universality in the opposite direction; (d) translation universals have been found to be genre-sensitive, for instance, while there is a tendency towards explicitation in German translation of popular scientific texts, this is not the case to the same degree for economic texts; (e) translations may be influenced by the status of the language

[4]Federico Zanettin's web page can be found at the following URL address: https://sites.google.com/site/federicozanettinnet/cl-htm#TOC-Translation-driven-Bilingual-and-Multilingual-Corpora.

[5]Available at: http://sslmitdev-online.sslmit.unibo.it/corpora/corporaproject.php?path=E.P.I.C.

of the source text genre, which in turn may influence the nature of the translation text genre and also the nature of comparable texts in the same genre (House 2008: 11). Becher (2011) shares House's critical stance. The departure point for his corpus study of English-German and German-English translations of business texts was not the assumption that explicitation is a translation-inherent universal process. Instead, Becher predicted that every instance of explicitation (and implicitation) can be accounted for by lexicogrammatical and/or pragmatic factors. His findings confirmed this hypothesis.

Malmkjær's (2008) contribution to the ongoing debate on the posited existence of translation universals is both critical and constructive. She suggests that universals such as simplification, explicitation and normalization would be better accounted for by the norm concept and explained on socio-cultural grounds. Instead, the Unique Items Hypothesis (UIH) is a good candidate for universal status because it can be explained on cognitive grounds. Indeed, the UIH, which has been confirmed by studies carried out with unrelated languages (Swedish and Danish on the one hand, and Finnish on the other), is a phenomenon that is not triggered by the source text, but seems to arise during the translation process, from the under-representation in a translator's mental lexicon of unique features of the target language. Malmkjær argues that, if the concept of the translation universal is to retain any theoretical credibility, it would have to be reserved to phenomena such as the UIH, "for which it makes sense to produce a cognitively based explanation" (Malmkjær 2008: 57). In addition to the quest for translation universals, other research projects were pursued during the 2003–2013 decade. They concern the style of literary translators, the role of ideology in determining translation choices and the study of Anglicisms, of which more later (see Laviosa 2011 for an overview).

In Applied Translation Studies corpora continued to be used to retrieve and examine lexical, terminological, phraseological, syntactic and stylistic equivalents. They also began to be utilized in Translation Quality Assessment (TQA) (Bowker 2003a, b). So, corpora have been increasingly incorporated in the curricular design of postgraduate translator training programmes to satisfy the exigencies of today's globalized and technologized language industry (Koby and Baer 2003; Zanettin et al. 2003; Kelly 2005; Ulrych 2005; Olohan 2007; Rodrigo 2008; Beeby et al. 2009). In sum, we can affirm that a coherent interdisciplinary theory combined with the professional and institutional recognition of corpora as valuable linguistic resources and translation aids has given rise to an effective partnership that is playing a crucial role in engendering a culture of research in education.

The question I wish to address at this point is the extent to which corpora have provided "an opportunity to reengage the theoretical and pragmatic branches of Translation Studies, branches which over and over again tend to disassociate, developing slippage and even gulfs", as was envisaged by Tymoczko (1998: 658). My view is that the relationship between corpus-based descriptive and applied studies has been open and reciprocal to a degree. Let me explain what I mean by this with two examples taken from the quest for translation universals. As we saw earlier, this line of enquiry was conceived as a descriptive research endeavour. Its findings were then projected into Applied Translation Studies where the Unique Items Hypothesis was tested and confirmed experimentally in the undergraduate translation classroom to raise awareness among students of what translation entails (Kujamäki 2004). Also, simplification and explicitation were tested as possible indicators of translation quality with a view to improving teaching methods and assessment criteria at postgraduate level. Simplification was found to correlate with lower-scoring translations and explicitation was found to correlate with higher-scoring ones (Scarpa 2006).

I believe that these studies, which engage in classroom-based investigations inspired by the insights provided by the pure branch of the discipline, represent the beginning of a new trend in Translation Studies. Moreover, I believe it is a promising orientation not only because it aims to replicate descriptive investigations and render translation teaching more effective and evaluation more rigorous, but also because it empowers students to gain a deep and critical understanding of the process, product and function of translation. Thanks to this knowledge, translator trainees will be capable of adhering to or innovating culturally-determined norms in an informed, conscious and responsible way. As Pekka Kujamäki contends, theories, models, concepts and experimentation with students should have an essential role in translation pedagogy "not only in research seminars but also and above all in the translation class: they open a way to novices' better understanding of their future status as experts of human translation" (Kujamäki 2004: 199).

In line with this envisioned direction for CTS, I propose that the holistic approach to translating culture elaborated by Tymoczko (2007) be adopted as a theoretical framework within which corpora can reengage the pure and applied branches of the discipline for the benefit of both of them. So, in the second part of my paper, I first expound the notion of holistic cultural translation and then I put forward the idea that corpora be used to foster this approach in two interrelated ways, i.e. through multilingual and multicultural research and pedagogy. The latter includes translation and language education as well as the training of translator trainers and language teachers.

1.4 Looking to the Future: Corpora and Holistic Cultural Translation

The holistic approach to translating cultural difference presupposes that translation be conceived as an open, cluster concept with blurred edges. This notion of translation, which Tymoczko calls the 'cluster concept translation' (or 'translation with an asterisk'), is defined in terms of resemblances between translation and three forms of cultural interface, i.e. representation, transmission and transculturation. I shall now define each of these large superordinate categories that partially encompass, impinge on and illuminate translation.

As a form of representation, translation offers an image or likeness of another thing, it stands in place of another entity and has authority to substitute for or act in place of that entity. Almost all translations are forms of representation, with a few exceptions such as pseudo translations (or fictitious translations) (Toury 1995/2012: 47–59). As a form of transmission, translation involves different types of transfer from one language and culture to another. Translations typically relay the content, language, function or form of the source text. The variability of methods adopted by translations that privilege transfer is very wide. It ranges from close textual fidelity to various degrees of manipulation of the linguistic features of the original. Many factors influence the vast array of transmission procedures adopted in translation practices, e.g. linguistic asymmetries, translation technologies, literacy practices, economic conditions, cultural sufficiency or enclosure, receptiveness to difference, aesthetic norms, taboos about certain types of content, asymmetries in power and cultural prestige as well as ideology (Tymoczko 2007: 119).

As a form of transculturation, translation is the transmission and uptake of borrowed cultural forms in the receptor environment and the consequent creation of new cultural phenomena. Transculturation generally includes such elements as verbal materials, religious beliefs and practices, social and political organization, artistic forms as well as aspects of material culture including technology and tools, agricultural and industrial practices, clothing, food, housing, transport and media (Tymoczko 2007: 120). In textual domains, transculturation often entails transposing elements of a literary system, e.g. poetics, genres, tale types. It also involves the uptake of the elements expressed in or carried by language such as discourses and world views (2007: 121). Lexical borrowing from a donor to a receptor language is a form of transculturation that occurs through language contact. Philip Durkin's historical linguistic study of loanwords in English (2014) shows how the rich variety of the English lexis reflects the vast number of words it has borrowed from languages as varied as Latin, Greek, Scandinavian, Celtic, French, Italian, Spanish, Russian Hebrew, Maori, Malay, Chinese, Hindi, Japanese and Yiddish. Also, the study of Anglicisms, a linguistic and socio-cultural phenomenon that has been investigated extensively through corpora, has attracted the interest of a growing number of scholars in recent years, particularly in Europe, where there is a need of reconciling the role of English as a European *lingua franca* with the EU commitment to cultural and linguistic diversity (Anderman and Rogers 2005; Furiassi et al. 2012; House 2012).

An insightful corpus-based study of translation as a form of transculturation is Ji's (2013) historical linguistic investigation of the introduction, assimilation and appropriation in early modern Chinese language and culture of the western key term *nation* in its modern sense of 'nation state'. The study covers five decades: 1840–1850, 1860–1870, 1880–1890, 1900–1910 and 1910–1920. The results show that during the 1840–1850 decade, the term *nation* was translated from English, French, Dutch, German and Italian mostly by words denoting the ethnic composition of a nation (i.e. *zhonglei*, *zulei*, *yizu*, *zuzhong*), while the word *guomin*, which referred to the people representing a sovereign state, was significantly less frequent. But in the subsequent decades this pattern changed. The most frequent equivalents of *nation* became *guo* and *bang*, which denoted a geopolitical entity. Less frequent were the words used to refer to the people of a given country (e.g. *min* and *guomin*).

The cluster concept translation rests on the assumption that language and culture are closely intertwined and “culture is the domain where human differences are most manifest” (Tymoczko 2007: 221). When communicating across cultural differences, as Tymoczko argues, it is not sufficient to approach the representation of culture in a linear, piecemeal fashion and resolve the problems incorporated in surface elements of the text one by one, sentence by sentence until the translation is complete (Tymoczko 2007: 233). What is needed instead is a holistic approach. As Tymoczko explains, “a holistic approach to translating culture will begin with the largest elements of cultural difference that separate the source culture and the target culture as a framework for coordinating the particular decisions about culture that occur as the text is actually transposed into the target language” (Tymoczko 2007: 235).

In order to help translators accomplish such a complex task, Tymoczko offers a partial repertory of cultural elements that might be taken into account as a guide for interpreting the source text and for determining the overall representation of culture in the target text. The inventory comprises:

- Signature concepts of a culture
- Key words
- Conceptual metaphors
- Discourses
- Cultural practices
- Cultural paradigms
- Overcodings
- Symbols.

I will now define each of these large cultural elements in turn and illustrate them with examples from various languages and genres. As we shall see, many of these examples are offered by corpus research. Signature concepts express key values in the social and economic organization of a culture. The words denoting them are highly connoted and rich in cultural associations. In early medieval Irish texts, for instance, words belonging to the semantic field of heroism, such as *honour*, *shame*, *taboo*, fall under the category of signature concepts (Tymoczko 2007). The signature concepts of contemporary American society can be equated to the values that American citizens cherish and are encouraged to promote. These are “hard work and honesty, courage and fair play, tolerance and curiosity, loyalty and patriotism”, as we read in the letter that the President of the United States of America sends to every new American citizen. On the other side of the Atlantic, the liberal values held by British people today are *openness*, *tolerance*, *compassion* and *strength*, as claimed by the Deputy Prime Minister, Nick Clegg in his speech delivered at the Liberal Democratic Party Conference held in York on 9 March 2014.

Key words are words that may point either to the signature concepts of a culture or to the thematic cultural elements chosen by a writer or speaker to structure a given text or a corpus of texts. For example, the strongest key words analyzed by Norman Fairclough (2000) in the corpus of New Labour texts (which contains a variety of texts produced under the New Labour Government led by the British Prime Minister Tony Blair from 1994 to 1999) are: *New Labour*, *new deal*, *new Britain*, *business and partnership*, *welfare reform* (Fairclough 2000: 17–20).

Drawing on the work of cognitive linguists such as Lakoff and Johnson (1980), conceptual metaphors shape the way we understand and experience reality, i.e. they structure (at least in part) what we do and how we understand what we are doing (Lakoff and Johnson 1980: 5). An example of variation in conceptual metaphors across languages is offered by Ding et al. (2010). Their corpus-based analysis of the metaphorical representation of the topic FEAR and its Chinese equivalent KONGJU, reveals that Chinese does not have the English conceptual metaphors FEAR IS A SUPERNATURAL BEING/A DISEASE/A SHARP OBJECT/A POISON/A LEGACY/A MACHINE. Moreover, the shared metaphor FEAR IS AN OPPONENT tends to be used in English to conceptualize the state of falling victim to fear, whereas in Chinese it is usually used to conceptualize an attempt to control it.

Ideological discourses are representations and visions of the social world expressed in speech and writing and as such they motivate action and cultural practice. They are the object of study of Critical Discourse Analysis, an area of research which has been investigated extensively through corpora. An example of such analysis is offered by Fairclough's research into the political discourse of the 'Third Way' in Tony Blair's speeches from 1998 to early 1999. The Third Way signifies a programme that was defined by centre and centre-left British governments as being neither old left nor 1980s right. It was built upon the notion of "the new global economy", that was accepted "as an inevitable and unquestionable fact of life" upon which politics and governments were to be premised (Fairclough 2000: 15,150). In the European Union the discourse of 'unity in diversity', which first came into use in 2000 "signifies how Europeans have come together, in the form of the EU, to work for peace and prosperity, while at the same time being enriched by the continent's many different cultures, traditions and languages."[6]

Cultural practices such as naming practices, forms of address and titles, the naming of kinship relationships play an important role in constructing personal and social identities and achieving social cohesion. They too may vary across languages. In English, for example, the word *grandfather* means 'father of one's father or mother' and the word *grandmother* means 'mother of one's father or mother'. The Italian equivalents are: *nonno* and *nonna* respectively. But in Thai the word ปู่ (po) means 'father of one's father', the word ตา (ta) means 'father of one's mother', the word ย่า (ya) means 'mother of one's father', and the word ยาย (yay) means 'mother of one's mother'. Similarly, in Swedish *farfar* = father's father, *morfar* = mother's father, *mormor* = mother's mother and *farmor* = father's

[6] http://europa.eu/about-eu/basic-information/symbols/motto/index_en.htm.

mother. In Chinese there are five equivalents of the English word *uncle*, i.e. *shushu*, *bobo*, *jiujiu*, *guzhang*, and *yizhang*, each referring to a specific family relationship.

Cultural paradigms pertain to humour, argumentation, logical sequencing in a text or the use of tropes. They tend to vary from language to language and within the same language over time. For example, a corpus-based study carried out by Niu and Hong (2010) on rhetorical repetition in English and Chinese print ads published in two leading newspapers in Singapore shows different patterns. The four most frequent repetition types in English are alliteration, rhyme, assonance, anaphora, while in Chinese they are assonance, anaphora, alliteration, rhyme. So, the results show that English uses more alliteration and rhyme than Chinese and Chinese uses more assonance and anaphora than English in this particular genre.

Overcodings are "linguistic patterns that are superimposed on the ordinary ranks of language to indicate a higher-order set of distinctions in language practices" (Tymoczko 2007: 243). They signal specific literary genres (e.g. poetry or narrative) and modes of communication (e.g. spoken or written). They also comprise rhetorical devices such as intertextuality, quotation and allusion. For example, the literary style of Latino writers in the United States is characterized by a constant code-switching from English to Spanish. As Díaz Pérez (2012: 171–172) observes, "[b]y introducing Spanish words, phrases or syntactic constructions into their English texts, they try to evoke the feeling of living on a *frontera*, of inhabiting two worlds which can be conflicting and complementary at the same time."

Within the category of overcodings we also find forms of textual structuring pertaining to aspects of register, dialects and languages for special purposes. An excellent example of corpus-based research that throws light into the relationship between overcodings and cultural context is Meng Ji's investigation of the lexico-grammatical features that characterize scientific language in early modern Chinese. This specialized register was developed by translating scientific texts from Western languages, most notably English, French and Dutch from the mid-nineteenth century to the turn of the twentieth century. This was a time characterized by the expansion of capitalism and imperialism in Asia. Two types of overcodings were unveiled by Ji's study: (a) dysillabic word structure, i.e. words created by combining two existing characters, and (b) functional particles.

Functional particles were created in Chinese to relay the meanings and functions expressed by the prefixes and suffixes of Latin and Greek origin that characterized Western scientific writing. Ji's study reveals two groups of functional particles, i.e. grammatically modified and semantic-cognitive functional particles. An example of the former is *de*, which identifies an adjective and was retrieved from ancient Chinese literary fiction. An example of the latter is *zhe*, an abstract term for things, agents or concepts, which was retrieved from ancient philosophical and historical texts as well as biographical essays. As Ji observes, while the original affixes "reflect the systematicity and continuity of the development of modern scientific language based on ancient Latin and Greek cultures and thoughts" (Ji 2012: 255), the development of equivalent functional particles in early Chinese scientific language "involved a thorough and painstaking re-examination of the target language body, searching for expressions of metaphorical references parallel to their Western counterparts" (Ji 2012: 255).

Finally, symbols are related to the identity of an individual, family, class, nation or deity (Tymoczko 2007: 145, fn. 28). Indeed, flower symbolism varies from language to language. Lilacs stand for light and early summer in Sweden but in Italy they represent envy. In some English villages a lilac branch may signify a broken engagement (Anderman 2007: 3). Folklore provides many other symbols and icons. In Indian mythology the word *naga* describes any kind of semi-divine serpent associated with water and fluid energy. *Nagas* are ambivalent deities, they are believed to bestow wealth and assure abundant crops but revoke these blessings if offended. An example of a symbol that crosses cultural boundaries is the poppy flower. During the years that followed the Great War it was adopted as the symbol of remembrance, especially in the Commonwealth countries. The idea was inspired by a poem written by the Canadian military doctor and artillery commander Major John McCrae in memory of his friend, a young Canadian artillery officer, Lieutenant Alexis Helmer, who was killed on 2nd May, 1915 when an exploding German artillery shell landed near him during the early days of the Second Battle of Ypres (Corni and Fimiani 2014: 307).

In Flanders Fields[7]

by John McCrae, May 1915

In Flanders fields the poppies blow
Between the crosses, row on row,
That mark our place; and in the sky
The larks, still bravely singing, fly
Scarce heard amid the guns below.

We are the Dead. Short days ago
We lived, felt dawn, saw sunset glow,
Loved and were loved, and now we lie
In Flanders fields.

Take up our quarrel with the foe:
To you from failing hands we throw
The torch; be yours to hold it high.
If ye break faith with us who die
We shall not sleep, though poppies grow
In Flanders fields.

[7]http://www.greatwar.co.uk/poems/john-mccrae-in-flanders-fields.htm.

At the 11th hour on the 11th day of the 11th month of 1918, Paris time, the Great War ended with the signing of the Armistice of Compiègne between the Allies and Germany.[8] The poppy flower is now the emblem of Remembrance Day (or Remembrance Sunday). This is the Sunday nearest 11th November in the UK and Canada when the country honours the people who died in the First and Second World Wars. Every year, on 11th November at 11 o'clock people in the UK observe a two-minute silence in memory of the fallen. On the same day, poppies, made of paper or plastic, are worn by veterans and citizens not only in Great Britain, but also in some European countries such as France and Belgium. (Corni and Fimiani 2014: 307).

As Tymoczko maintains, considering all the above cultural elements helps translators compare their own culture with the source culture as it is reflected in texts. In order to make these cross-cultural comparisons translators need to develop self-reflexivity. It is through self-reflexivity that they will be able to identify those elements of cultural difference that need to be mediated. As a result, "a holistic approach to cultural translation rather than a selective focus on a limited range of cultural elements enables greater cultural interchange and more effective cultural assertion in translation, allowing more newness to enter the world" (Tymoczko 2007: 233). And, I wish to add, corpora can play an important role in fostering holistic cultural translation since they can fruitfully be used "to illuminate both similarity and difference and to investigate in a manageable form the particulars of language-specific phenomena of many different languages and cultures" (Tymoczko 1998: 657).

1.5 Towards a Corpus-Based Holistic Pedagogy

How can corpora be used to unearth cross-cultural differences and similarities in research as in practice? I think an effective way of achieving this goal is to work towards a multilingual pedagogy that espouses the tenets of holistic cultural translation and incorporates corpora not only as tools for acquiring technical skills, but also as resources for developing translingual and transcultural competences that enable translators to act as self-reflexive, responsible meaning makers in our increasingly globalized, multicultural world. I recommend that in such envisioned pedagogy comparable and parallel corpora be explored through discovery and justification procedures (Toury 1995/2012) so as to infer culturally determined norms and particularities on the basis of empirical evidence.

What follows is an example of how this teaching method was adopted to investigate a form of transculturation, i.e. the relationship between the loanword

[8]http://www.history.com/this-day-in-history/world-war-i-comes-to-an-end.

business in Italian and its English etymon. Loanwords are problematic in translation as their lexico-grammatical profiles tend to be different across donor and receptor languages. An Anglicism may, for example, convey only a subset of the senses expressed by the English etymon. Also, an Anglicism may acquire different connotations in the receiving language. As Pulcini (2002: 162) explains, lexical borrowing is a complex phenomenon "because it involves referential, connotative, contextual and sociocultural components of meaning". Consequently, the senses conveyed by words in the donor language may be "kept, altered, restricted or expanded" in the receiving language (Pulcini 2008: 196). Normally, if an English word is borrowed in order to fill a semantic gap in Italian, the referential meaning remains the same, as is the case with the terms *agribusiness* or *bed and breakfast.* But in many other instances, changes tend to occur in the form of restriction or expansion.

An example of restriction is offered by the term *benchmark*. In Italian it refers to a financial market index that enables investors to assess the upward or downward trend of an investment fund (Pulcini 2008: 197). In English *benchmark* is used as a verb and a noun. As a verb it means "to provide a standard that something can be judged by". As a noun it means "an amount, level, standard etc. that you can use for judging how good or bad other things are" (Macmillan English Dictionary for Advanced Learners 2007). An example of expansion of meaning and change of word class is provided by the borrowed term *backstage*. As an English adverb it means "in the area behind the stage in a theatre, including the rooms where the actors get dressed". As an adjective it means "relating to the area behind the stage in a theatre, including the rooms where the actor get dressed: *a backstage pass* (= a special ticket that allows you to go backstage)" (Macmillan English Dictionary for Advanced Learners 2007). In Italian the noun *backstage* refers to the area behind the stage in a theatre and to a documentary that illustrates the technical problems, atmosphere, emotions and gossip involved in the preparation of a film, event or theatre performance (Pulcini 2008: 198).

These cross-linguistic asymmetries largely arise from the fact that "meanings are established in individual languages by contrasts of similar items in semantic fields" (Görlach 2003: 93). Hence, translator trainees working out of English often find it difficult to decide when to use Anglicisms appropriately. Since translated texts can serve as semantic mirrors reflecting meaning across languages (Johansson 2003: 136), they are an invaluable resource for investigating loanwords across donor and receptor languages. Before moving on to report on a small-scale research project conducted in the postgraduate translation classroom to raise awareness of translation norms and the phenomenon of lexical borrowing, I will outline the procedural steps we adopted in keeping with the methodology elaborated by Toury (1995/2012) for discovering regularities in translational behaviour.

1.6 Toury's Discovery and Justification Procedures

The research methodology proposed by Toury (1995/2012: 31–34) for descriptive translation studies is articulated in three phases. The first phase starts with the selection of individual translations or a corpus of translations within the target culture. Toury's perspective is target-oriented: translations are texts that belong to the target culture, they are texts in their own right, not just mere representations of their source texts. The analysis carried out in the first phase involves the initial assessment of the acceptability of the individual translations or corpus of translated texts without reference to the source texts. Acceptability is the extent to which a translated text adheres to the linguistic and cultural norms prevailing in the target language for a particular text genre. The opposite concept is that of adequacy. An adequate translation is one which leans towards the norms of the source language and culture and contains traces of the textual features of the source text.

The second phase starts with the identification of the source texts and proceeds to comparing the target texts and their sources in parallel, that is sentence by sentence, paragraph by paragraph. The aim is to determine target-source relationships, translation problems, translation solutions and shifts. According to Toury they can be of two kinds: obligatory, which are caused by systemic differences between the source and target languages, and non-obligatory, which are motivated by literary, stylistic or cultural considerations.

In the third phase of the analysis, the relationships established between the target texts and their sources become the basis of first-level generalizations about the initial norm underlying the concrete way in which equivalence is realized. The initial norm governs the basic choice which can be made between adequacy (which involves adhering to source norms) and acceptability (which involves subscribing to norms originating in the target culture). Adequacy and acceptability are to be considered as two poles of a continuum where the target text can be positioned on the basis of its linguistic features examined vis-à-vis the source text and comparable original texts produced in the target language (Toury 1995/2012: 79–85). Equivalence is not conceived as an a priori notion that is based on an absolute criterion of adherence to the source text. This means that in a descriptive study the researcher will always assume that equivalence exists. What s/he will unveil is the actual way in which it is realized in terms of the balance between invariance and transformation. This type of equivalence in turn constitutes a stepping stone for discovering the concept of translation that informs the target texts examined, this being defined in terms of the acceptability-adequacy continuum.

At each stage of this process of gradual discovery of facts about the nature of translation and translating, hypotheses are formulated on the basis of empirical descriptions and then verified through further procedures that are applied to an expanding corpus to achieve higher and higher levels of generalization. The procedures elaborated by Toury are largely compatible with the Data-Driven Learning approach developed by Johns (1991a, b) in foreign language education. This method, as we discussed earlier, is also employed in corpus-based translator

Table 1.1 Procedural steps for the analysis of an ever expanding KWIC concordance

Step 1: Initiate	Look at the words that occur immediately to the right of the node word to note any that are repeated; do the same with the words to the left of the node and decide on the strongest pattern
Step 2: Interpret	Look at the repeated words to form a hypothesis that may link them
Step 3: Consolidate	Look for other evidence that can support the hypothesis formulated in Step 2
Step 4: Report	Write out the hypothesis formulated in Step 2 and revised according to the evidence collected in Step 3 so as to have an explicit, testable version
Step 5: Recycle	Start with the next most important pattern near the node going through the same steps as before, and then look for the strongest pattern remaining on either side, until there are no repeated patterns
Step 6: Result	Make a final list of hypotheses linking them in a final report on the node word
Step 7: Repeat	Gather a new selection of concordances and apply your report on this new data, going through the same steps and confirming, extending or revising the list of hypotheses drawn up in Step 6

training, particularly in the teaching of Languages for Special Purpose (LSPs). The notion of 'discovery' plays an important role in Toury's and Johns' methodologies. They both require that students and researchers alike progress from empirical data to generalization. The basic corpus-based procedure adopted by Johns is to "Identify—Classify—Generalise" (Johns 1991a: 4) the lexico-grammatical features associated with words that are particularly problematic for advanced learners. The main tool for carrying out this analysis consists of KWIC concordance lines and the procedural steps are those proposed by Sinclair (2003: xvi–xvii) (Table 1.1):

Case study I, in Sect. 7, is an illustration of how Toury's and Johns' methods of enquiry together with Sinclair's conceptual definitions were integrated in a corpus-based methodology devised for the student-centred, professionally-oriented translation classroom.

1.7 Case Study I: Translating *Business* in Italian

A corpus- and translation-driven investigation was performed by the students attending a 60-credit postgraduate course in specialized translation at the University of Bari Aldo Moro during the 2008–2009 academic year. As part of a 3-credit module devoted to the Language of Business and Economics, I designed a teaching unit on Anglicisms. The learning objectives were as follows: (a) to become familiar with corpora as one of the computer-aided translation tools and resources available to the professional translator; (b) to discover the textual-linguistic norm underlying the translation of polysemic lexical Anglicisms; (c) to discover Italian native equivalents of well-established English loanwords.

The sources of data consist of: (a) a corpus of 71 translated and non-translated comparable articles taken from the Italian weekly magazine *Economy* and (b) a corpus of 71 English articles from *The Economist* and their translations in *Economy*. In the first phase, the lexical Anglicisms contained in the translational subcorpus were identified. By Anglicism we intend "a word or idiom that is recognizably English in its form (spelling, pronunciation, morphology or at least one the three), but is accepted as an item of the vocabulary of the receptor language" (Görlach 2003: 1). The most frequent Anglicism was *business*, a well-established English loanword, having been introduced in the Italian lexicon in 1895, as attested by the Vocabolario della lingua italiana di Nicola Zingarelli (2004).

Out of nearly 60,000 running words, 37 occurrences of *business* were retrieved. The analysis of the KWIC concordance lines revealed five discrete meanings of *business*.

I. the work of producing or buying and selling goods or services for money.
II. a high profile area of business where more than one company operates.
III. a. a highly profitable business activity undertaken by a company; b. investment, deal or transaction made by a company.
IV. a large organization that provides services, or that makes or sells goods.
V. volume of business.

For each of the above meanings the collocation, colligation, semantic preference and semantic prosody were identified using the definitions provided by Sinclair (2003: 173–178):

1. identify collocational profile (lexical realizations), i.e. two or more words occurring near each other in a text;
2. identify colligational patterns (lexico-grammatical realizations), i.e. the occurrence of a grammatical class or structural pattern with another one, or with a word or phrase;
3. consider common semantic field (semantic preference), i.e. a clear preference in the structure of a phrase for words of a particular meaning;
4. consider pragmatic realizations (semantic prosody), i.e. the special meaning conveyed by words grouping together, which relates not so much to their dictionary meanings as to the reasons why they were chosen together; it has been recognized in part as connotation, pragmatic meaning and attitudinal meaning.

The results of the students' analysis were as follows:

I. *Business* occurs with words that refer to other human activities, (*turismo e business*), the geographical place where business is carried out and the people of different nationalities that are in business; it forms multi-word-units (*business hub*, *area business*, *segmento business*, *aree consumer e business*).

II. *Business* occurs with nouns that identify a particular business sector and the position gained in the market, nouns referring to the major players that operate in or impact on it, adjectives describing its qualities such as diversity, profitability or importance.
III. *Business* occurs with words that refer to the company undertaking a particular business activity and to the type of activity undertaken, verbs referring to the changes undergone by a business, adjectives and nouns describing its main features such as novelty, solidity or volatility; it forms one compound (*core business*).
IV. *Business* occurs with words referring to the people owning or running a company or to the way in which a company organizes its activities; it forms multi-word-units (*business model*, *modello di business*, *business manager*, *business partner*).
V. *Business* occurs with words that refer to the monetary value (or turnover) of a company or business sector:

Al 73enne Ecclestone è rimasto il 25 % del gruppo che gestisce il business da 800 milioni di dollari.

[73 year old Ecclestone still owns 25 % of the group that runs the 800 million dollar business].

Per contrastare il cambiamento del clima anche gli ambientalisti riscoprono il nucleare. Un business da 125 miliardi di dollari.

[To counteract climate change, even environmentalists are rediscovering nuclear power].

All the five senses conveyed by *business* did not show a clear tendency to represent either unpleasant or pleasant states of affairs. Hence, the semantic prosody of *business* was considered to be neutral. Next, *business* was examined in the comparable subcorpus of non-translated articles. The number of occurrences was nearly double, 74 against 34. The meanings identified were the same, except for senses II and V, which also referred to illegal business activities. The analysis of the KWIC concordance lines revealed sameness and difference (examples of the latter are highlighted in bold):

I. *Business* occurs with words that refer to other spheres of human activity (*business e società*, *sport e business*, *musica e business, business e genetica*), the geographical place where business is carried out and the people of different nationalities that are in business; it forms multi-word-units (*business information*, *aree di business*, *clienti business, utenti business*). ***Business*** **is used in creative collocations:** ***Il business resta in porto*** [business stays in the port]; ***Titanic del business*** [the Titanic of business]; ***Il business non è l'unico quadrante su cui far girare le lancette della vita*** [Business isn't the only dial on which the hands of life turn].

II. *Business* occurs with nouns that identify a particular sector and the position gained in the market, nouns referring to the major players that operate in or impact on it, adjectives describing its qualities such as diversity, profitability or importance. ***Business*** **forms compounds:** ***business travel***, ***social business, business online***. ***Business*** **refers to illegal sectors**: ***il business dei falsi*** [the business of counterfeits]; ***i business si chiamano droga, prostituzione, racket*** [drugs, prostitution and racket are known as businesses]. ***Business*** **is used in creative collocations**: ***un business che si chiama sconto*** [a business that is called discount]; ***un business duro come il teak*** [a business as hard as teak].
III. *Business* occurs with words that refer to the company undertaking a particular business activity and the type of activity undertaken, verbs referring to the changes undergone by a business activity, adjectives and nouns describing its main features such as novelty **as well as importance, competitiveness, credibility or profitability; it forms various compounds**: ***core business***, ***business continuity, business case, business plan***. **It strongly collocates with:** ***possibilità, opportunità, occasioni, fare***. **It is used in creative collocations and puns**: ***il business in una cannuccia*** [business in a straw]***; ora faccio business col cuore*** [now I'm doing business with my heart]; ***un Tornado di business*** [a business Tornado]; ***ho più di un business per capello*** [I've got more than one business in my hair].
IV. *Business* occurs with words referring to the people owning or running a company or to the way in which a company organizes its activities; it forms multi-word-units (*business unit, unità di business*, *business development manager*). ***Business*** **is used in creative collocations and puns**: ***un business fatto di nuvole*** [a business made of clouds]; ***l'Enav e quel business che è caduto dal cielo*** [Enav and the business that fell from the sky]; ***il business lievita alla luce del sole*** [business rises in the sunlight].
V. *Business* occurs with words that refer to the monetary value (or turnover) of a company or business sector. **It is used in creative collocations**: ***Quel business da 2,5 milioni di sacchi di caffè*** [That business worth 2.5 million sacks of coffee]. **It refers to illegal activities**: ***Il business [dei falsi] vale almeno 7 miliardi di euro all'anno*** [the business of counterfeits is worth 7 billion euros a year]; ***Cibo Nostro. La Mafia nell'alimentare***. ***Quasi 20 miliardi di incassi per la criminalità organizzata***: ***tanto vale oggi il business mafioso nell'agroalimentare, nelle sue varie declinazioni*** [Our Food. The Mafia in the food sector. Almost 20 billion euros worth of takings for organised crime: that's how much the Mafia business is worth in the agricolture and food sectors in its various ramifications].

The results show that the collocation, semantic preference, colligation and semantic prosody of *business* in translational and non-translational Italian appear to be divergently similar. Moreover, translators seem to have resisted the influence of English by limiting the use of *business*. The next phase involved mapping the

Italian target texts onto the English source texts. For each of its five meanings, the following native Italian equivalents of *business* were retrieved:

(I) *il mondo degli affari*, *gli affari*, *affari*, *l'attività*, *attività commerciali*
(II) *il settore*, *l'industria*, *le industrie*
(III) *un'attività commerciale*, *l'attività*, *le attività delle aziende*
(IV) *un'azienda*
(V) *generating the business* → *cedendo i prestiti*.

The textual-linguistic norm that was inferred from these findings is a preference for native Italian equivalents. There is also one example (see meaning V above) where the original non-finite verb phrase, *generating business*, was translated with an equivalent expression, *cedendo i prestiti* (relinquishing loans), which explicates the original sense. Interestingly, the preference for 'domestic competitors' (Görlach 2003) of *business* in translational Italian is consistent with Musacchio's (2005:76) study of a parallel/comparable corpus of economics articles, which shows a lower percentage of lexical borrowings in translational versus non-translational business Italian.

The final phase of the analysis was carried out with the subcorpus of original English articles vis-à-vis the subcorpus of comparable non-translated Italian articles. It involved identifying the lexico-grammatical shifts in the use of *business* across the donor and the receptor language. Unlike in English, *business* appeared to have pejorative overtones in Italian when it conveyed meanings II and V. For these two senses there was, therefore, some evidence of a negative semantic prosody. Also, while in English *business* was found to refer to a small or medium enterprise, in Italian it was found to denote a large company (meaning IV). Finally, in Italian *business* was sometimes used in creative expressions that usually appeared in titles or subtitles and fulfilled the pragmatic function of attention-getting devices.

A further comparison was carried out between the results obtained in class through corpus-based analysis and the description of *business* in a sample of monolingual and bilingual dictionaries. This exercise revealed the special contribution of corpora as resources for gaining a deeper understanding of language use in the target language. Here is the entry for *business* in the *Grande dizionario italiano dell'uso* (GRADIT) (1999–2000):

> **business** /ˈbIzIns/(bu-si-ness) s.m.inv. ES ingl. [1895; ingl. *Business*/ˈbIzns/, der. di *busy* "affaccendato"] attività economica, spec. di grande rilievo, molto redditizia: *entrare in un grande b.*, *il b. delle telecomunicazioni* | anche con riferimenti a rilevanti attività illegali: *il b. della droga*, *gli appalti sono un grande b.*

GRADIT describes *business* as an economic activity, especially a high profile and very lucrative one. The examples reported in the dictionary are: "entering a large business", "the telecommunications business". The entry also states that *business* may refer to illegal activities, e.g. "the drug business", "calling for tenders is big business". Students noted that in GRADIT senses II and III had been conflated and the negative semantic prosody that was found to be associated with *business* was recorded only for sense II. The other senses they had unearthed in their corpus-based investigation were not recorded.

The case study presented here focuses on an authentic problem raised by the students themselves, namely the norm governing the translation of polysemic Anglicisms in a specific subject domain and text type, i.e. the language of business and economics in periodicals articles. Students' evaluations at the end of the course revealed to me that their learning experience aided by corpora and inspired by a holistic pedagogy was perceived as 'professionally empowering' (Kiraly 2003). Indeed, they stated that learning through corpora had equipped them with the specialized knowledge, self-reliance, authentic experience and expertise they need to acquire in order to become language professionals with highly developed intercultural communicative competences. Moreover, as a teacher, I felt professionally empowered too, because I learnt from my students as much as I taught them about translation and language for specific purposes. Also, I was able to engage in what Kiraly (2003) calls 'action research', which is carried out by teachers in their own classroom with a view to bringing about long-term changes in pedagogy.

1.8 Case Study II: Translating *RIBA* in Italian

In her monograph, *Exploring Corpora for ESP Learning*, Laura Gavioli (2005) puts forward a corpus-based methodology for learning English for special purposes that draws on corpus linguistic theory and integrates contrastive analysis and translation in the method's design and procedures. The "search and discovery" approach elaborated by Johns (1991a, b), the "learner as a traveller" principle proposed by Bernardini (2000) and John Swales' "the learner as a spy" metaphor (1990) provide the foundations of Gavioli's pedagogy. This is exemplified by an array of learning tasks she carried out with undergraduate students at the University of Modena and Reggio Emilia, Italy over a period of ten years. What follows is an illustration of how Gavioli utilized the proposed teaching methodology in the translation classroom.

The problems that students often encounter when dealing with languages for special purposes can be grouped under the category of overcodings. As we saw earlier, these include forms of textual structuring pertaining to aspects of register, dialects and languages for special purposes (Tymoczko 2007). A feature of textual structuring in specialized registers is the use of acronyms. ESP learners often find it difficult to understand their exact meaning not only because they are not always recorded in dictionaries and other reference material such as encyclopedias, but also because of students' lack of subject-specific knowledge. An example in medical English is the acronym *RIBA*, which stands for "recombinant immunoblot assay".

As Gavioli (2005) recounts, her students encountered this acronym for the first time when translating a research article on hepatitis C from English to Italian. The problem they faced was not so much to find a direct equivalent in Italian since they knew from past experience that English acronyms are generally borrowed in Italian, but to understand its conceptual meaning, i.e. the values and conventions shared by the medical discourse community that uses this particular term. Their lack of subject-specific knowledge constituted a major obstacle for understanding the whole text. To solve this problem, students investigated, together with their teacher, a small corpus of 12 research papers on hepatitis C, which they had previously compiled for teaching/learning purposes.

A total of nine KWIC concordance lines were retrieved for the node word *RIBA* (the full set of concordances is reproduced below from Gavioli (2005: 92–93)). Lines 2 and 4 contain the acronym's component words, but the students were unable to make sense of them. The examination of the other concordances revealed the co-occurrence of *RIBA* with *confirmatory tests* (lines 2, 3, 9) and *confirmed* (line 7). Also, *RIBA* was found in the vicinity of the expressions *supplementary tests* (line 1), *were then tested* (line 1) and *were also tested* (line 5). This finding suggested to the students that *RIBA* was an additional test that is probably performed to confirm results obtained from other tests. Moreover, the analysis revealed that *RIBA* is a test performed on blood serum (lines 1, 4, 5, 7, 8, 9) it is similar to another test called *LIA* (lines 1, 8, 9) and it is used to confirm hepatitis C virus infection as well as confirm the results from other tests, namely *ELISA* (lines 4, 7, 9) and *PCR* (line 3). Students also discovered that one of the most likely reasons why *RIBA* is used as a confirmatory test rather than as part of large scale screening is its high cost (line 6).

(1) identified as antibody positive. Enough serum was available from 42 patients to carry out 2 supplementary tests, a LIA (Innogenetics) and a RIBA (Chiron). The results obtained by all the tests are shown in Table II. All 4 antibody detection systems yielded consistent results in 28

(2) d in duplicate. As confirmatory tests, the INNO-line immunoassay (LIA) (Innogenetics) and a second-generation recombinant immunoblot assay (RIBA) (Chiron) were used. HBV Markers Hepatitis B surface and e antigen and their antibodies were measured using ELISA (Abbot

(3) n 6 patients. Of the 68 patients, 15 were antibody negative using both ELISAs. In 8 (54 %) viral RNA was detected by PCR (Table 1.1). LIA and RIBA could be performed in only 2 of the cases (Table II). In one case both confirmatory tests yielded positive results, thus confirming

(4) considered the anti-HCV titer. Certain serum samples which were positive in ELISA-II were then tested with the recombinant immunoblot assay (RIBA) (Chiron, Emeryville and Ortho, Raritan, NJ, U.S.A.). This assay uses all 4 recombinant HCV antigens (C-100; 5-1-1; C-33c

(5) 3 and capsid polypeptides; selected cases were also tested against synthetic peptides derived from different regions of the HCV sequence. PCR and RIBA II immunoblot assays were done on selected sera. Each of 55 probable and 5 of 11 possible hepatitis cases who were seronegative before transfusion

(6) ral causes of transaminase elevation by our clinical review committee may also have played a role. The multiple epitope immunoblot assay (RIBA II) is an attractive solution to the problem of multivalent testing; however, this assay is too expensive for large scale research use and is

(7) of anti-HCV positive cases than ELISA-I in both resolved (21/22, 95 %) and chronic cases (91/92, 100 %). ELISA-II results were confirmed by RIBA in 48 out of 51 (94.1 %) cases tested, and were indeterminate in the remaining 3 (5.9 %). When certain sera which had been obtained

(8) own in Fig. 3, HCV RNA was detected by PCR in 4 different serum samples drawn previously from this patient. In 3 patients not tested by LIA and RIBA, serum samples drawn at earlier times were investigated and in 2 of these patients, virus was already detectable in all of the earlier samples

(9) the second generation ELISA clearly had a higher detection rate. To control the reliability of the ELISA results, 2 confirmatory tests (LIA and RIBA) were carried out with a part of the serum samples. However, clear discrepancies between the 4 antibody assays were observed in 33 % of the cases

Through group work, students were able to discuss their interpretation of corpus-based data and became acquainted with a number of issues and problems facing the hepatitis C medical community. These entail the diagnosis of the virus as well as the use and evaluation of many different tests and their relative costs. The problem-driven and inductive procedures that students adopted in their search for meaning involved formulating hypotheses and checking them by consulting reference works such as encyclopedia entries and medical handbooks. As Swales (1990) would put it, students were successful in acting as spies of the medical discourse community, discovering not only the propositional meaning of its specialized lexis, but also its lexico-grammar and the conventions which characterize the world of knowledge it signifies.

In addition to the investigation of ESP corpus data and reference material, students compared the English concordances with those retrieved from a comparable Italian corpus. Their goal was to fully understand the meaning of *RIBA* and its conventional usage in English as well as become aware of the conventions in their native language. Their investigation of a total of seven concordances retrieved from a small corpus of hepatitis research articles in Italian—see full set reproduced below (Gavioli 2005: 104)—made it very clear to them that *RIBA* is a confirmatory test, as shown by the use of the verb forms *conferma* (lines 1, 2), *confermata* (line 3), *confermati* (line 4), *confermato* (line 5).

1. pazienti presentavano un test di conferma in RIBA indeterminato. Si trattava di una
2. ‘HCVAb (determinazione e/o conferma in RIBA); d) “C + pregressa B”
3. positività per anti-HCV, confermata con RIBA, nelle epatobiopsie con segni istologici
4. positive all’ELISA sono stati confermati al RIBA, tranne 2, che sono risultati indeterminati
5. HCV EIA) e nei positivi confermato con RIBA di II generazione (Ortho-Chiron
6. on RIBA di II generazione (Ortho-Chiron Riba HCV), sul siero di 88 pazienti
7. poi sui sieri conservati, lo screening in RIBA-Ortho. I controlli ematologici

1.9 Conclusion

The agenda I wish to set out for corpus studies of translation now and in the foreseeable future is to stimulate a continuous process of mutual exchange between theory, research and practice so as to reengage the pure and applied branches of the discipline. This entails undertaking research that is firmly grounded in theory and is driven by concerns arising in practice. It also means promoting practice that is underpinned by the insights given by research and elaborating theory that is substantiated and refined by empirical findings. This mutual exchange will, I think, promote the interaction between academics and professionals (Pym 2001) and go a long way to resolving the traditional dichotomy between 'the academic and the vocational in translator education' (Kearns 2008). For those who are interested in pursuing this scholarly endeavour, I am pleased to announce the launching of *Translation and Translanguaging in Multilingual Contexts* (TTMC), published by John Benjamins. This new journal provides a forum for interdisciplinary research that aims to create new synergies within and beyond Translation Studies.

References

Anderman, Gunilla (ed.). 2007. Introduction. In *Voices in Translation. Bridging Cultural Divides*, 1–5. Clevedon: Multilingual Matters.

Anderman, Gunilla, and Margaret Rogers (eds.). 2005. *In and out of English: For better, for worse?*. Clevedon: Multilingual Matters.

Beeby, Allison, Patricia Rodríguez Inés, and Pilar Sánchez-Gijón (eds.). 2009. *Corpus use and translating*. Amsterdam: John Benjamins.

Becher, Viktor. 2011. Explication and implicitation in translation: A corpus-based study of English-German and German-English business texts. Unpublished PhD Dissertation. University of Hamburg. http://ediss.sub.uni-hamburg.de/volltexte/2011/5321/pdf/Dissertation.pdf. Accessed 18 Nov 2014.

Bernardini, Silvia. 2000. *Competence, capacity, corpora: A study in corpus-aided language learning*. Bologna: CLUEB.

Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In *Inter-lingual and inter-cultural communication: Discourse and cognition in translation and second language acquisition studies*, eds. Juliane House and Shoshana Blum-Kulka, 17–35. Tübingen: Gunter Narr.

Blum-Kulka, Shoshana, and Eddie A. Levenston. 1983. Universals of lexical simplification. In *Strategies in interlanguage communication*, ed. Klaus Faerch, and Gabriele Kasper, 119–139. London: Longman.

Bowker, Lynne. 2003a. Towards a collaborative approach to corpus building in the translation classroom. In *Beyond the ivory tower: rethinking translation pedagogy*, eds. Brian James Baer and Geoffrey S. Koby, 193–210. Amsterdam: John Benjamins.

Bowker, Lynne. 2003b. Corpus-based applications for translator training: Exploring the possibilities. In *Corpus-based approaches to contrastive linguistics and translation studies*, ed. Sylviane Granger, Jacques Lerot, and Stephanie Petch-Tyson, 169–183. Amsterdam: Rodopi.

Bruti, Silvia, and Pavesi Maria. 2008. Interjections in translated Italian: Looking for traces of dubbed language. In *Investigating English with Corpora: Studies in Honour of Maria Teresa Prat*, ed. Aurelia Martelli, and Virginia Pulcini, 207–222. Monza: Polimetrica.

Chesterman, Andrew. 1998. Causes, translations, effects. *Target* 12(1): 200–230.

Díaz Pérez, Francisco. 2012. Writing and translating on the Frontera: Language and identity in Liliana Valenzuela's translations of Latina writers in the United States. In *Traducción y (A)culturación en la era global/translation and (Ac)culturation in the global era*, ed. Catalina Iliescu Gheorghiu, 171–180. Alicante: Editorial Agua Clara.

Ding, Yan, Dirk Noël, and Hans-Georg Wolf. 2010. Patterns in metaphor translation: Translating fear metaphors between English and Chinese. In *Using corpora in contrastive and translation studies*, ed. Richard Xiao, 40–61. Cambridge: Cambridge Scholars Publishing.

Durkin, Philip. 2014. Borrowed words. *A history of loanwords in English*. Oxford: Oxford University Press.

Espunya, Anna. 2007. Is explicitation in translation cognitively related to linguistic explicitness? A study of interclausal relationships. In *The study of language and translation. Special issue of Belgian journal of linguistics*, eds. Willy Vandeweghe, Sonia Vandepitte and Marc Van de Velde, 21: 67–86.

Fairclough, Norman. 2000. *New labour, new language?*. London: Routledge.

Furiassi, Cristiano, Virginia Pulcini, and Félix Rodríguez Gonzáles (eds.). 2012. *The anglicization of the European lexis*. Amsterdam: John Benjamins.

Gaspari, Federico, and Silvia Bernardini. 2010. Comparing non-native and translated language: Monolingual comparable corpora with a twist. In *Using corpora in contrastive and translation studies*, ed. Richard Xiao, 215–234. Newcastle upon Tyne: Cambridge Scholars Publishing.

Gavioli, Laura. 2005. *Exploring corpora for ESP learning*. Amesterdam: John Benjamins.

Görlach, Manfred. 2003. *English words abroad*. Amsterdam: John Benjamins.

Grande dizionario italiano dell'uso (GRADIT). 1999–2000. Turin: UTET.

Granger, Sylviane, Jacques Lerot, and Stephanie Petch-Tyson (eds.). 2003. *Corpus-based approaches to contrastive linguistics and translation studies*. Amsterdam: Rodopi.

Gustavo, Corni, and Enzo Fimiani (eds.). 2014. *Dizionario della Grande Guerra: Cronologia, Stati, personaggi, eventi, eserciti, simboli, culture, eredità*. Textus Edizioni: L'Aquila.

House, Juliane. 2008. Beyond intervention: Universals in translation. Trans-kom 1(1): 6–19. http://www.trans-kom.eu/bd01nr01/transkom_01_01_02_House_Beyond_Intervention.20080707.pdf.

House, Juliane. 2012. Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages. In *Corpus-based translation studies: Research and applications*, ed. Alet Kruger, Kim Wallmach, and Jeremy Munday, 187–208. London: Continuum.

Ji, Meng. 2012. Translation and scientific terminology. In *Quantitative methods in corpus-based translation studies*, ed. Michael P. Oakes, and Meng Ji, 251–273. Amsterdam: John Benjamins.

Ji, Meng. 2013. Conflicts and interactions in early modern chinese scientific translations. In *Translation, history and arts. New horizons in Asian interdisciplinary research*, eds. Meng Ji and Atsuko Ukai, 1–26. Newcastle upon Tyne: Cambridge Scholars Publishing.

Johansson, Stig. 2003. Reflections on corpora and their uses in cross-linguistic research. In Corpora in translator education, eds. Federico Zanettin, Silvia Bernardini, and D. Stewart, 135–144.

Johns, Tim. 1991a. Should you be persuaded: Two examples of data-driven learning. In *Classroom concordancing*, eds. Tim Johns and Philip King. Special Issue. English Language Research Journal 4: 1–16. Manchester: St. Jerome.

Johns, Tim 1991b. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In *Classroom Concordancing*, eds. Tim Johns and Philip King. Special Issue. English Language Research Journal 4: 27–46.

Kearns, John (ed.). 2008. The academic and the vocational in translator education. In *Translator and interpreter training. Issues, methods and debates*. 184–214. London: Continuum.

Kelly, Dorothy. 2005. *A handbook for translator trainers*. Manchester: St. Jerome.
Kiraly, Donald C. 2000. *A social constructivist approach to translator education. Empowerment from Theory to Practice.* Manchester: St. Jerome.
Kiraly, Donald C. 2003. From instruction to collaborative construction: A passing fad or the promise of a paradigm shift in translator education? In *Beyond the ivory tower: Rethinking translation pedagogy*, eds. Brian James Baer and Geoffrey S. Koby, 3–27. Amsterdam: John Benjamins.
Koby, Geoffrey S. and Brian James Baer (eds.). 2003. Task-based instruction and the new technology. Training translators for the modern language industry. In *Beyond the ivory tower: Rethinking translation pedagogy*. 211–227. Amsterdam: John Benjamins.
Kruger, Alet, Kim Wallmach, and Jeremy Munday (eds.). 2011. *Corpus-based translation studies: Research and applications*. London: Continuum.
Kujamäki, Pekka. 2004. What happens to "unique items" in learners' translations? "theories" and "concepts" as a challenge for novices' views on "good" translation. In *Translation universals: Do they exist?*, ed. Anna Mauranen, and Pekka Kujamäki, 187–204. Amsterdam: John Benjamins.
Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.
Laviosa, Sara. 2011. Corpus-based translation studies: Where does it come from? Where is it going? In *Corpus-based translation studies: Research and applications*, ed. Alet Kruger, Kim Wallmach, and Jeremy Munday, 13–75. London and New York: Continuum.
Macmillan English Dictionary for Advanced Learners. 2007. London: Macmillan.
Malmkjær, Kirsten. 2008. Norms and nature in translation studies. In *Incorporating corpora. The linguist and the translator*, eds. Gunilla Anderman and Margaret Rogers, 49–59. Clevedon: Multilingual Matters.
Mauranen, Anna, and Pekka Kujamäki (eds.). 2004. *Translation universals: Do they exist?*. Amsterdam: John Benjamins.
Musacchio, Maria Teresa. 2005. The influence of English on Italian: The case of translations of economics articles. In *In and out of English: For better, for worse?*, eds. Gunilla Anderman and Margaret Rogers, 71–96. Clevedon: Multilingual Matters.
Niu, Guiling, and Huaqing Hong. 2010. Repetition patterns of rhetoric features in English and Chinese advertisements: A corpus-based contrastive studies. In *Using Corpora in Contrastive and Translation Studies*, ed. Richard Xiao, 433–456. Cambridge: Cambridge Scholars Publishing.
Olohan, Maeve. 2007. Economic trends and developments in the translation industry. *The Interpreter and Translator Trainer* 1(1): 37–63.
Pulcini, Virginia. 2002. Italian. In *English in Europe*, ed. Manfred Görlach, 151–167. Oxford: Oxford University Press.
Pulcini, Virginia. 2008. Corpora and lexicography: The case of a dictionary of anglicisms. In *Investigating English with corpora: Studies in honour of Maria Teresa Prat*, eds. Aurelia Martelli and Virginia Pulcini, 189–203. Monza: Polimetrica.
Pym, Anthony. 2001. To localize and humanize … On academics and translation. *Language International* 13(4): 26–28.
Rodrigo, Elia Yuste (ed.). 2008. *Topics in language resources for translation and localisation*. Amsterdam: John Benjamins.
Russo, Mariachiara, Claudio Bendazzoli, and Annalisa Sandrelli. 2006. Looking for lexical patterns in a trilingual corpus of source and intepreted speeches: Extended analysis of EPIC (european parliament interpreting corpus). *Forum* 4(1): 221–254.
Scarpa, Federica. 2006. Corpus-based specialist-translation quality assessment: A study using parallel and comparable corpora in English and Italian. In *Insights into specialised translation*, ed. Maurizio Gotti, and Susan Šarčević, 155–172. Bern: Peter Lang.
Setton, Robin. 2011. Corpus-based interpreting studies. In *Corpus-based translation studies: Research and applications*, ed. Alet Kruger, Kim Wallmach, and Jeremy Munday, 33–75. London: Continuum.

Shlesinger, Miriam. 1998. Corpus-based interpreting studies as an offshoot of corpus-based translation studies. Meta 43(4): 486–93. http://www.erudit.org/revue/meta/1998/v43/n4/. Accessed 18 Nov 2014.

Shlesinger, Miriam. 2009. Towards a definition of interpretese. An intermodal, corpus-based study. In *Efforts and models in interpreting and translation research. A tribute to Daniel Gile*, eds. Gyde Hansen, Andrew Chesterman, and Heidrun Gerzymisch-Arbogast, 237–253. Amsterdam: John Benjamins.

Sinclair, John McH. 2003. *Reading concordances*. London: Pearson Education.

Steiner, Erich. 2008. Explicitation—towards an empirical and corpus-based methodology. In *Meaning in context: Implementing intelligent applications of language studies*, ed. Jonathan J. Webster, 235–278. London: Continuum.

Steiner, Erich. 2005. Explicitation, its lexicogrammatical realization, and its determining (independent) variables—towards an empirical and corpus-based methodology. *SPRIK reports. Reports of the project languages in contrast* (*Språk i kontrast*) 36. http://www.hf.uio.no/ilos/forskning/prosjekter/sprik/pdf/Report_36_ESteiner.pdf. Accessed 18 Nov 2014.

Swales, John. 1990. *Genre analysis*. Cambridge: Cambridge University Press.

Tirkkonen-Condit, Sonia. 2004. Unique items—over- or under-represended in translated language? In *Translation universals. Do they exist*?, eds. Anna Mauranen and Pekka Kujamäki, 177–186. Amsterdam: John Benjamins.

Toury, Gideon. 1995/2012. Descriptive translation studies and beyond. Amsterdam: John Benjamins.

Tymoczko, Maria. 1998. Computerized Corpora and the future of translation studies. Meta 43(4): 652–60. http://www.erudit.org/revue/meta/1998/v43/n4/. Accessed 18 Nov 2014.

Tymoczko, Maria. 2007. *Enlarging translation, empowering translators*. Manchester: St. Jerome.

Ulrych, Margherita. 2005. Training translators: Programmes, curricula, practices. In *Training for the new millennium: Pedagogies for translation and interpreting*, ed. Martha Tennent, 3–33. Amsterdam: John Benjamins.

Ulrych, Margherita, and Simona Anselmi. 2008. Towards a corpus-based distinction between language-specific and universal features of mediated discourse. In *Investigating English with corpora: Studies in honour of Maria Teresa Prat*, ed. Aurelia Martelli, and Virginia Pulcini, 257–273. Monza: Polimetrica.

Vocabolario della Lingua italiana di Nicola Zingarelli. 2004. *Undicesima Edizione*. Bologna: Zanichelli.

Xiao, Richard (ed.). 2010. *Using corpora in contrastive and translation studies*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Xiao, Richard, Lianzhen He, and Ming Hue. 2010. In pursuit of the “third code”. Using the ZJU corpus of translational Chinese in translation studies. In *Using corpora in contrastive and translation studies*, ed. Richard Xiao, 182–214. Newcastle upon Tyne: Cambridge Scholars Publishing.

Zanettin, Federico. 2012. *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Manchester: St. Jerome.

Zanettin, Federico, Silvia Bernardini, and Dominic Stewart (eds.). 2003. *Corpora in translator education*. Manchester: St. Jerome.

Zuffery, Sandrine, and Bruno Cartoni. 2014. A multifactorial analysis of explicitation in translation. *Target* 26(3): 361–384.

Chapter 2
History in Keywords

2.1 Introduction

In her presentation about the 10-year history of corpus-based translation studies, Laviosa (2003) divided this period into three stages. The first one, labelled by her as "the dawn", comprises the years 1993–1996, when the idea of using corpus methods for descriptive translation research was introduced by Baker (1993, 1995, 1996). The second stage covers the time period starting from 1996, when Laviosa-Braithwaite (1996) published her Ph.D. thesis, and continuing until the end of the decade. These years were characterised by an intensive study on translation universals (e.g. Olohan 2000; Mauranen and Kujamäki 2004). During the third stage, including the first years of the new millennium, the range of the research field expanded. From the perspective of the present study it is interesting that Laviosa mentioned as one of the new fields *corpus-based critical translation studies*. This term was used by her for corpus-based translation studies that have common ground with critical linguistics and critical discourse analysis.

Now that another ten years have elapsed it is apparent that the interest towards translation universals did not fade away in the beginning of the new millennium. Conversely, as Zanettin (2012: 23) puts it, "[t]he search for regularities in corpora of translations has been mostly carried out with the aim of investigating universals." He adds that some studies with more specific aims have been carried out as well, and mentions two scholars—Williams and the author of the present study. Williams (2009) examines Spanish translations of the verb 'report' in biomedical research articles. Kemppanen's (2004, 2008) studies concern ideology-bound lexis in Russian-Finnish translations compared to original Finnish texts on the political history of Finland.

The present study continues work with ideological features of lexis. The aim of the study is twofold: first, to introduce different methods for corpus-based research with an orientation towards the historical study of lexical features of translations,

S. Laviosa et al., *Textual and Contextual Analysis in Empirical Translation Studies*, New Frontiers in Translation Studies,
DOI 10.1007/978-981-10-1969-2_2

and, second, to discuss how the corpus-based approach can be combined with qualitative methods (cf. Laviosa 2002: 89).

The corpus analysis focuses on describing ideology-specific features in different sub-corpora by using corpus linguistic applications—keywords, key-keywords, word clusters and concordances. In addition, the paper provides an example of using corpus methods together with a qualitative analysis of narrative structure. The narrative analysis is based on Greimas's (1980) model of narrative structures, which is applied for analysing contexts of certain ideology-bound keywords.

The material of this study comprises a Finnish-language corpus referred to as *The Comparable Corpus of History Texts.*[1] It consists of 0.5 million words of texts translated from Russian into Finnish and 0.5 million words of non-translated texts (non-fiction books on political history). The corpus of translated texts consists of different sub-corpora reflecting different circumstances of text production: texts from the Soviet Era (282,938 words), the Perestroika Era (38,830 words) and the post-Soviet Era (193,313 words). The analysis focuses on the two bigger corpora—the Soviet Era translations dating from the 1970s and the post-Soviet translations from the 1990s. The reference corpus consists of non-translated history texts (537,154 words) written in Finnish by Finnish historians and published between 1971 and 1999.

2.2 Studying Ideology-Bound Lexis

2.2.1 Ideology as a Research Object

The starting point of the present study is the assumption that every text fulfils a certain *ideological function* (Kemppanen 2005: 187–188). This assumption is based on the ideas of Vološinov 1972 ([1930]: 18), also cited by Petrilli (1992: 252), on the existence of an ideological function attached to every verbal sign. From the perspective of quantitative study on representations of ideology, this kind of approach appears problematic. What is the point in counting manifestations of a phenomenon, if it is said to be present everywhere? In addition, as Vološinov (1972 [1930: 18]) states, the same linguistic items of a language can be used for verbalising any ideology. Every verbal sign gets its actual ideological meaning in the social context where it is produced and received by the language users. Still, it has been claimed that texts differ from each other in the explicitness of representing their ideological features (Fairclough 1997: 26). A political speech expresses values laden in the text itself more clearly than, for example, a text book in physics. The

[1]*The Comparable Corpus of History Texts*. Compiled at the University of Joensuu (nowadays the University of Eastern Finland) within the research project *Translated Finnish and Translation Universals. A Corpus-based Study* (1997–2001) led by Prof. Anna Mauranen.

latter—maybe considered “ideologically neutral” by most readers—supports the power of science and technical culture just by being compulsory reading in schools.

Despite the fuzziness of the concept of ideology, some scholars have carried out studies to identify its markers. These studies come from the area of critical discourse analysis—both in linguistics and translation studies. Among them could be mentioned studies carried out by the following authors: Fowler and Kress (1979), Thompson (1994 [1990]), Fairclough (1995), Knowles and Malmkjær (1989), Hatim and Mason (1997) and Puurtinen (1998).

The markers of ideology differ in their applicability to corpus-based study. The most suitable issues for quantitative research are e.g. tautology in a text (cf. Sowarka 1987) or translators' solutions concerning repetition (cf. Knowles and Malmkjær 1989). Hunston (2002: 109) stresses that ideological stances can be identified in texts with several methods: analysis of keywords, lexical patterning, semantic prosody and repetitions.

However, it must be added that corpus linguistic analysis does not deal just with the use of disconnected words. A multidimensional study of texts including the study of collocations, word clusters and keywords in larger contexts makes it possible to get a clearer view of manifestations of such a complex phenomenon as ideology. Some scholars emphasise the existence of two levels of observations. Hatim (1999, cited in Laviosa 2000: 160) stresses that corpus-based analysis enables us to tap into both what is *in the text* and what is *from the text*. ‘In the text’ means the language itself, which can be analysed by text analysis. ‘Of the text’ refers to the text as a whole, including its overall influence on the level of ideology. Baker (1999: 282) emphasises the importance of viewing the results from the micro-level analysis in a social and ideological context. She sees focusing on formal structures and ignoring the social and ideological context as one of the reasons for a negative attitude towards the role of linguistics in translation studies in the 1980s and 1990s. The same idea of combining different approaches has been expressed by Olohan (2004: 149) as well. She sees it fruitful to link together corpus data and sociocultural studies by relating the results of corpus analysis to the production and reception of translations.

2.2.2 Keywords, Keyness and Ideology

The present study introduces how *keyword analysis* can be used for identifying ideology-bound lexical features. The keyword method requires the use of a comparable corpus. In our case the material consists of translated and non-translated texts on the political history of Finland.

Scholars in corpus studies have different views on the focal issue of *keyness*. Scott and Tribble (2006: 55) state that “many languages use the metaphor *key* to identify people, places words, ideas as important; the term is used in text databases for retrieval and in practice it does not seem to be defined. Keyness is, therefore, a quality which is intuitively obvious.” Scott and Tribble stress that we should talk about keyness as a primarily textual phenomenon suggesting what is important in this concrete text (or a set of texts), what the text is all about. Scott and Tribble stay at the textual level, but they mention some attempts by other researchers to go further and make claims about keyness on the cultural level. Williams (1976) compiled a dictionary of keywords in British culture. This contribution was based on intuitive considerations about the keyness of English words. Stubbs (1996, 2001) applied this idea to corpus-based identification of cultural keywords. Actually, the idea of defining keyness on the cultural level is included already in Firth's (1969 [1957]) concept of *focal* or *pivotal words* referring to 'sociologically important words'. He used this term in his article "The technique of semantics", written as early as 1935. It is interesting that Firth (1968 [1957]: 180) refers to this concept also in his article "A synopsis in linguistic theory, 1930–1935" and uses the term *key words* as a synonym for *focal words*. Firth's *technique of semantics* can be considered as the model for current corpus-based studies of keywords. This project on textual analysis included both a formal and sociological component. Firth developed a method for defining the contextual distribution of words. He emphasises that words should be analysed in their actual contexts (Firth 1969 [1957]: 10; Stubbs 1996: 165–166).

Applying corpus methods to monolingual comparable corpora enables us to study manifestations of ideology. They can be analysed as proportional linguistic features of two corpora. The main notion of our study, *keywords*, is well suited for this purpose. A list of keywords reflects a relationship between an analysed text (or a set of texts) and a reference corpus. As Scott (1998: 65) puts it: "Keywords are those whose frequency is unusually high in comparison with some norm". The word *norm* refers here to a frequency norm, which is used for defining the quantitative features of the text corpus in focus. When examining translations, the researcher usually takes the position that non-translated texts stand for the "normal", "typical" or "domestic". For example, corpus studies on translation universals, the so called T-universals, deal with notions such as simplification, untypical lexical patterning and under-representation of target-language-specific items (Chesterman 2011: 177). All of these suggested universals characterise translations as something "untypical" compared to the norm of non-translated texts.

When analysing translations and non-translations from the point of view of ideology, it must be taken into account that the juxtaposition of "typical" and "untypical" is also present here. The concept of ideology has been approached by many scholars from an evaluative perspective. Ideological meanings have been understood as something negative (Fairclough 1997: 66–67) and the term *ideology* has been used for characterising only the ideology of others. Referring to Marx and Engels, this viewpoint of the concept has been labelled as *Die deutsche Ideologie* (Kellner 1998: 67–74). Using quantitative markers of ideology, like corpus-driven lexical features in our study, makes it possible to avoid a one-sided position in defining what is "ideological" in a text. In translation research this means that the analysis can be focused on lexical features of non-translated texts produced in the ideological framework of the target culture as well. Comparison of statistical information on language use in translated and non-translated texts may reveal regularities in non-translated texts that might remain unnoticed in research with an orientation only towards “untypicalities” of translations.

2.3 Combined Methods: Keywords and Narrative Analysis

This section presents a method which combines two different approaches: a corpus-based, quantitative analysis and an *actantial analysis*, a qualitative analysis of narrative structures in the texts. The analysis starts from generating a keyword list from the corpus of translations by using the corpus program *WordSmith Tools* (Scott 2014). The list functions as a basis for further analysis. The next step consists of comparing the keywords in corpora representing different historical periods. After this, keywords are examined in their contexts: firstly, by analysing frequent word clusters and secondly keywords in narrative structures.

2.3.1 Keyword List as a Starting Point

A multi-method corpus analysis starts with retrieving a keyword list from a comparable corpus. The list used in this study indicates the keywords in translated history texts. The list illustrated below (Table 2.1) actually shows the so called *key-keywords* (Scott 1998)—the words that are key in more than one text (labelled in the table as *key files*).[2] The key-keywords tool excludes from a list words that characterise a certain text, but are rare in other texts. For example, the frequency of a certain personal or geographical name can be high in one book, where this name is relevant, but in a set of texts representing the whole genre this lexical item may show very low frequency. The words appearing repeatedly as keywords in several texts are of greater interest from the perspective of possible markers of ideology.

Table 2.1 presents facts about the corpus of translated history texts. The words in this list can be classified into several semantic groups. The words *Suomi* ‘Finland’, *Neuvostoliitto* ‘Soviet Union’, *Venäjä* ‘Russia’ and *Leningrad* are related with each other by their geographical or political character. The first three are used for naming the two neighbouring countries, which were closely linked through foreign policy. *Leningrad* refers to a big city in the Western part of the Soviet Union/Russia. In the context of political history this proper name has a “priming

[2]This key-key-word list originates from an earlier study (Kemppanen 2008), where the lemmatised version was in use. The remainder of the present study is based on the non-lemmatised version of the corpus.

Table 2.1 Keywords in translations[a]

Position	Word	Number of "key files"	Number of "key files" %
1	Suomi 'Finland'	6	54.55
2	ja 'and'	6	54.55
3	Neuvostoliitto 'Soviet Union'	5	45.45
4	minä 'I'	3	27.27
5	tuo 'that'	3	27.27
6	vuosi 'year'	3	27.27
7	Leningrad	3	27.27
8	tuollainen 'that kind of'	2	18.18
9	hän 'he/she'	2	18.18
10	sota 'war'	2	18.18
11	sopimus 'agreement, treaty'	2	18.18
12	suhde 'relationship, relations'	2	18.18
13	hallitus 'government'	2	18.18
14	Venäjä 'Russia'	2	18.18
15	yhteistyö 'co-operation'	2	18.18
16	välinen 'between'	2	18.18
17	ystävyys 'friendship'	2	18.18

[a]Keyness has been calculated via log-likelihood

effect" that makes the reader expect information about the strategic importance of the city. In addition, the word *hallitus* 'government' could be placed in this word group. In a political text this lexical item can be used for referring to a state that the text is about: *The government/the government of Finland/Finland does not allow something*.

The second category can be built around the semantic field 'relationship between actors'. This group includes three words with a meaning of a positive relationship (*sopimus* 'agreement, treaty', *yhteistyö* 'co-operation' and *ystävyys* 'friendship') and one referring to a negative relationship (*sota* 'war'). In addition, the word *suhde* 'relationship, relations' itself belongs to the group. It can be associated with either a positive or a negative relation to something or somebody.

The third group comprises the following words: *tuollainen* 'that kind of', *välinen* 'between', *minä* 'I', *hän* 'he, she', *tuo* 'that' and *ja* 'and'. These lexical items differ from those mentioned above as far as their semantic function is concerned. They all are used as function words (i.e. grammatical words), while the words in the previous categories pertain to content words (semantic words). At first glance function words may appear neutral from the point of view of verbalising ideology. Still, their contexts reveal, that these words can be used for building ideology-bound constructions together with the content words as well.

The only word in the list that was left outside the mentioned categories is the word *vuosi* 'year'. It is of course logical that this word is typical in history texts because of their time-orientation, but it is surprising that when two sub-corpora of history texts are compared the word is still in the list. Why is it much more common in translations? A look at the concordances of the word *vuosi* 'year' shows that one of the main reasons for the difference in frequencies between the comparable corpora lies in the common use of the phrase *vuoden* [genitive of *'vuosi'*] *1948 sopimus* 'the agreement of the year 1948' in translated texts (192 occurrences). This item is a lexical calque of the Russian expression *Dogovor 1948 goda* appearing only twice in non-translated texts. In Finnish this document is labelled with the acronym expression *YYA-sopimus* 'The Agreement of Friendship, Cooperation, and Mutual Assistance', also known as the *YYA Treaty*. This phrase occurs in non-translated texts 130 times, which implies the importance of this concept in the foreign policy between Finland and the Soviet Union, regardless of how it is named. The YYA Treaty has been considered as one of the most obvious signs of the Soviet influence on Finnish foreign policy.

2.3.2 Historical View on Keywords

The next step of the analysis is to look closer at the different layers of time in the key word list above. The corpus of translated texts is categorised into three blocks according to the year of publication. Five of the eleven texts represent the Soviet Era, two of them the time of transition, Perestroika, and four texts belong to the post-Soviet time. This study concentrates on comparing the texts produced before Perestroika and after the collapse of the Soviet Union.

Table 2.2 introduces a list of keywords which has been made of the five translations from the Soviet Era. Unlike in the previous section, a non-lemmatised version of the corpus is used, so in some cases several word forms of the same

Table 2.2 Keywords in the Soviet Era corpus

	Key word	Freq.	%	RC. Freq.	RC. %	Keyness
1	*Suomen* 'Finland', gen.[a]	4929	2.25	5203	0.97	1765.42
2	*ja* 'and'	9992	4.56	15,723	2.92	1204.47
3	*Neuvostoliiton* 'Soviet Union', gen.	1965	0.90	2010	0.37	740.44
4	*sopimuksen* 'agreement, treaty', gen.	481	0.22	288	0.05	372.87
5	*Paasikivi* (proper name), nom.	467	0.21	274	0.05	369.28
6	*Neuvosto-Venäjän* 'Soviet Russia', gen.	308	0.14	125	0.02	328.94
7	*kanssa* 'with'	1258	0.57	1538	0.29	323.78
8	*rauhan* 'peace', gen.	370	0.17	237	0.04	267.58
9	*suhteiden* 'relations', gen. pl.	365	0.17	242	0.05	247.46
10	*ystävyyden* 'friendship'	149	0.07	24		246.62

[a]The case of each word form is indicated in the table. Plural indicated with *pl.*, singular not marked

lexical item appear in the list. When the material represents the lexis in an agglutinative language like Finnish, non-lemmatised material provides a lot of information about morphological features in language use. The table indicates the frequency of each key word in the corpus of the Soviet Era material. The first column on the left provides the frequency in absolute numbers, the second one shows the proportion of each word form as a percentage, the third shows frequency in the reference corpus of non-translated history texts, the fourth shows the proportion in the reference corpus and the last column gives the so called *keyness value.* It reflects how strongly the frequency of a word in the study corpus differs from its frequency in the reference corpus. The bigger the value, the bigger is the difference in frequencies between the comparable materials. The *KeyWord*-function generates a word list demonstrating the words with the highest keyness values. The corpus program automatically carries out the comparison of the frequency lists in focus, cross-tabulates the frequencies and performs the statistical tests. Word frequencies with maximal proportional difference in the comparable corpora draw a word into the keyword list (Scott 1998: 71).

Table 2.2 demonstrates the top of the keyword list extracted from the Soviet Era corpus. The total number of keywords generated amounted to 488 items, but even this short sample from the list gives a clear picture what the books on history in this material are about. The main characters in the story are the acting states—*Finland and the Soviet Union* (*Soviet Russia*). They have undoubtedly good *relations with* each other: they have made a mutual *agreement* and live in *friendship* and *peace*. One of the main architects of this foreign policy was *Paasikivi*, the President of Finland. This discourse is familiar to all who are acquainted with the clichés of the foreign policy between the Soviet Union and Finland. This policy was at its peak during the 1970s, precisely the time the translated books were published. The analysis of the keywords shows that although Finland has been accused of “Finlandisation”—for practising a foreign policy that flatters the policy pursued by the Soviet Union, there is still a difference in Finnish and Soviet discourse when verbalising these ideas in history writing. If the discourses had been alike, there would not have been so many keywords with this “friendship prosody” in the list.

Most of the words in the list (8 of 10) are semantic words, which stresses the differences in the sub-corpora concerning the content. The two grammatical words *ja* ‘and’ and *kanssa* ‘with’ match with the semantics of the content words referring to positive relations between actors. It is obvious that these grammatical items can be used for combining the names of the actors and positive-laden content words into longer clusters with clearly represented ideological meaning. An extract from the concordance (Figure 2.1) gives an idea how the grammatical word *kanssa* ‘with’ functions as a part of a larger lexical element with positive content: *rauhanneuvotteluista Neuvostoliiton kanssa* ‘peace negotiations with the Soviet Union’, *diplomaattista dialogia Neuvostoliiton kanssa* ‘diplomatic dialogue with the Soviet Union’, *suhteiden normalisointia Neuvostoliiton kanssa* ‘normalisation of relations with the Soviet Union’, *ystävyys ja yhteistyö Neuvoliiton kanssa* ‘friendship and co-operation with the Soviet Union’, *Yya-sopimus Neuvostoliiton kanssa* ‘the Agreement of Friendship, Cooperation and Mutual Assistance with the

Figure 2.1 Concordance of the word *kanssa* 'with'

Soviet Union', *hyvä naapuruus ja yhteistyö Suomen kanssa* 'good-neighbourliness and co-operation with Finland', *aktiiviseen yhteistyöhön Suomen kanssa* 'active co-operation with Finland', *rauhan ja yhteistyön suhteet Suomen kanssa* 'relations of peace and co-operation with Finland' *allekirjoitettiin rauhansopimus Suomen kanssa* 'peace agreement with Finland was signed', *hedelmällistä yhteistyötä Suomen kanssa* 'fruitful co-operation with Finland'.

The keyword list in the Table 2.3 looks very different to the previous one. It illustrates the material of translated history texts representing the post-Soviet Era.

The keyword list shown in the table does not give any picture of the basic theme or themes of the texts. However, it suggests a different genre or at least another perspective to the narrated content. This can be inferred from the range of pronouns and verb forms. Most of them are verbalisations of the first person singular. The lemma *minä* 'I' is represented in the list by three different case forms *minulle* 'to me, for me (allative)', *minua* 'me, of me (partitive)' and *minut* 'me (accusative)'. This paradigm continues with verb forms of the first person singular: *olen* 'I am', *olin*

Table 2.3 Keywords in the post-Soviet corpus

	Keyword	Freq.	%	RC. Freq.	RC. %	Keyness
1	*ja* 'and'	10,868	3.85	15,723	2.92	497.09
2	*minulle* 'to me, for me', allative	339	0.12	48		474.03
3	*en* 'not', negative, 1st pers. sg.	391	0.14	81	0.02	470.29
4	*minua* 'me', partitive	246	0.09	22		391.54
5	*sanoi* 'he/she said'	452	0.16	163	0.03	391.14
6	*olen* 'I am'	296	0.10	59	0.01	362.22
7	*olin* 'I was'	243	0.09	28		326.58
8	*ntenä*, ending used with certain numerals, 'on the 4th, 5th etc. of'	153	0.05	0		326.58
9	*minut* 'me', accusative	180	0.06	19		274.86
10	on 'he/she is'	2543	0.90	3123	0.58	267.22

'I was' and *en* 'not (negative in 1st pers. sg.)'.[3] All the listed word forms suggest that the texts on history in this sub-corpus favour a narrative in first person singular. Also over-presentation of the word *sanoi* 'he/she said' refers to claims of single human actors which are missing in the Soviet Era materials, where the focus is on the state level, manifested in the names of acting states and the president of the state.

A special case in the list is the abbreviation *ntenä* that is connected with a colon to certain ordinal numbers, e.g. *4:ntenä* 'on the 4th of' and *5:ntenä* 'on the 5th of'. It is clear that this keyword does not belong to the same group of markers as those mentioned above. This item reflects clearly the influence of a lexical pattern in the source texts and can be traced to the Russian way of indicating dates: *4-ogo oktjabrja* 'on the 4th of October'. The Finnish construction is a calque of the Russian one. This construction is possible in Finnish, but as the keyword list shows, it does not occur at least in this field of language use. The fact that the ending *ntenä* is totally missing in the reference corpus confirms a result from earlier studies (Kemppanen 2008) showing the over-representation of the names of the months in Russian-Finnish translations of history texts. The Finnish convention of expressing dates in history books is based on using numbers and dots, e.g. *4.10. 1941*, instead of attaching a suffix to the number and writing the name of the months with letters.

The fact that a keyword can be recognised as a clear manifestation of interference instead of being identified as an ideology-bound element distinguishing the two comparable corpora of texts is present in all keyword research with an orientation on the analysis of ideology in texts. The researcher is repeatedly forced to ask the question: "Is this ideology or interference? Must this lexical feature be interpreted as influence on the cultural level or purely on the linguistic level?" (cf. Kemppanen 2006).

The only keyword that is common to both lists is the word *ja* 'and' at the top of the lists. As was stated earlier, the key words tell what the text is about. This "aboutness" comes mainly from the semantic picture transferred by the content words in a keyword list. However, in many cases it may appear even more interesting to find out the reasons why a grammatical word has achieved a position in a list. The coordinating conjunction *ja* 'and' is a fascinating case, because it is one of the most frequent words in any Finnish text. Why does it appear as a keyword in history texts translated from Russian, and both in Soviet and post-Soviet translations? The earlier considerations on the semantic matching of the item *ja* 'and' to the Soviet discourse now appear in a new light. This question will be analysed more closely in the next section.

---

[3]In the Finnish language the negative word is conjugated for person. The verb form appearing with the negative remains unconjugated.

2.3.3 Keywords in Context: Clusters

The next step in this keyword study consists of a more detailed analysis of the contexts where a certain keyword has been used. As the main tools for more advanced analyses of words, corpus researchers mention mostly the terms *collocations*, *clusters* and *lexical patterning* (Olohan 2004: 148–149; Zanettin 2012: 130–146). Other terms have also been used. Jantunen (2004: 19) states that *collocation* refers mostly to the co-occurrence of two words and the term is often distinguished from *multi-word strings* (Mauranen 2000: 120), *lexical bundles* (Biber and Conrad 1999), *multi-word items* (Nelson 2000) and *clusters* (Kenny 2000: 99), used for labelling sections of three or more words. All these issues refer to a tendency of words to appear repeatedly together. Firth defines *collocations* as “actual words in habitual company” (Firth 1968 [1957]: 182). Sinclair (1991: 170) puts this more clearly by stating that a *collocation* is a tendency of two or more words to occur close each other. From the perspective of the present study it is worth mentioning that the relevance of lexical patterning has been recognised by scholars in critical linguistics. Heikkinen (1999: 163) stresses the role of language patterns in verbalising ideological meanings. According to him, the patterns arise repeatedly in a certain context as a result of so called *collocation schemes* where one element forces the other to keep company with the first one.

The patterns building up around the keywords show their *semantic prosody*. Louw (1993: 157) defines *semantic prosody* as “a consistent aura of meaning with which a form is imbued by its collocates”. According to Sinclair (1998: 20), semantic prosody is the reason why a language user selects a certain lexical item. The item is attached with positive or negative evaluation. A particular prosody that occurs recurrently in a text or a set of texts manifests a system of values—an ideology.

Examining collocations and multi-word strings allows us to identify lexical elements which manifest an ideology. Stubbs (1996: 169) considers the study of lexical patterns very important in investigating the problem of language and ideology. He claims that “fixed and semi-fixed expressions (collocations, catch phrases, clichés and idioms) encode cultural information”. Empirical evidence based on this kind of analysis provides information about “how the culture is expressed in lexical patterns”. The importance of large-scale context searches is stressed when the research object deals with culture-bound issues, such as ideology.

The idea of a multi-step keyword analysis is to examine the lexical features of a certain set of texts from various viewpoints. The study of the context where the keywords in focus appear can be carried out by analysing word clusters. The next two tables given below illustrate the most frequent 3-word clusters including the keyword *ja* ‘and’ in the two corpora of translated history texts: the texts from the Soviet Era and those from the post-Soviet period. The keyword *ja* ‘and’ was chosen as an object for further study, because it occurs as a keyword in both corpora. The analysis will show the apparent differences between the two corpora.

Table 2.4 Clusters with the word *ja* and' in the Soviet Era translations

	The cluster and the case(s) of the conjugated words	Number of clusters
1	*Neuvostoliiton ja Suomen* 'Soviet Union and Finland', gen.	707
2	*ja Suomen välisen* 'between… and Finland', gen./gen sg.	135
3	*Suomen ja Neuvostoliiton* 'Finland and the Soviet Union', gen.	134
4	*ja Suomen välisten* 'between… and Finland', gen./gen. pl.	90
5	*ja Suomen välisissä* 'between… and Finland', gen., inessive pl.	68
6	*että Neuvostoliiton ja* 'that… the Soviet Union and', gen.	52
7	*Neuvostoliitto ja Suomi* 'the Soviet Union and Finland', nom.	48
8	*Suomen ja Venäjän* 'Finland and Russia', gen.	48
9	*ja Suomen väliset* 'between… and Finland', gen./nom. pl.	45
10	*Venäjän ja Suomen* 'Russia and Finland', gen.	41
11	*rauhan ja turvallisuuden* 'peace and security', gen.	40
12	*ystävyys yhteistyö ja* 'friendship, co-operation and', nom.	40
13	*ja keskinäisestä avunannosta* 'and mutual assistance', elative	37
14	*Englannin ja Ranskan* 'England and France', gen.	33
15	*ja Suomen välinen* 'between… and Finland', gen.	31
16	*Neuvosto-Venäjän ja Suomen* 'Soviet-Russia and Finland', gen.	31
17	*ja Suomen välillä* 'between… and Finland', gen.	30
18	*ystävyyden ja yhteistyön* 'friendship and co-operation', gen.	30
19	*ja keskinäisen avunannon* 'and mutual assistance', gen.	28
20	*Repolan ja Porajärven* 'Repola and Porajärvi', gen. (geographical names)	28
	Number of clusters in total	1696

The list of clusters shown in Table 2.4 gives a clear picture of the discourse in the Soviet Era corpus. It repeats the lexical set provided in the list of individual keywords. The coordinating conjunction *ja* 'and' links, firstly, names of states co-operating for achieving common aims—e.g. *the Soviet Union* and *Finland,* and, secondly, nouns referring to positive action, such as *friendship, co-operation, mutual assistance*, *peace* and *security*. Some of the clusters do not fully show the co-operating partners or the coordinated actions, but the cluster elements give a hint of the missing part, e.g. *ja Suomen välillä* 'between … and Finland'. In the context of the other clusters the story as whole can be guessed. It is clear that in many cases the keyword *suhteet* 'relations' is somewhere outside the borders of the cluster. Even those clusters with the element 'between' can be seen as implicit markers of "the ideology of friendship". The last cluster in the list *Repolan and Porajärven* consists of two geographical names combined with the word *ja* 'and'. This string of words refers to a region that was offered by the Soviet Union as a compensation for its territorial claims elsewhere in Finland. These geographical names are thus politically value-laden as well.

The list of clusters including the word *ja* 'and' retrieved from corpus of the post-Soviet translations (Table 2.5) has some common features with the list of

Table 2.5 Clusters with the word *ja* and' in the post-Soviet Era translations

	The cluster and the case(s) of the conjugated words	Number of clusters
1	*Neuvostoliiton ja Suomen* 'the Soviet Union and Finland', gen.	135
2	*Suomen ja Neuvostoliiton* 'Finland and the Soviet Union', gen.	106
3	*Englannin ja Ranskan* 'England and France', gen.	55
4	*Venäjän ja Suomen* 'Russia and Finland', gen.	40
5	*Suomen ja Venäjän* 'Finland and Russia', gen.	33
6	*ja Suomen välisten* 'between … and Finland', gen./gen. pl.	31
7	*ja sanoi että* 'and said that'	28
8	*ja että se* 'and that it'	22
9	*ja sen jälkeen* 'and after that'	22
10	*Saksan ja Neuvostoliiton* 'Germany and the Soviet Union', gen.	21
11	*ja niin edelleen* 'and so on'	18
12	*Englanti ja Ranska* 'England and France', nom.	18
13	*ja ennen kaikkea* 'and first of all'	17
14	*Neuvostoliittoa vastaan ja* 'against the Soviet Union and', partitive	16
15	*ja lisäsi että* 'and added that'	14
16	*ja Suomen välisen* 'between … and Finland', gen./gen. pl.	14
17	*ja Neuvostoliiton välillä* 'between … and the Soviet Union', gen.	13
18	*ja ennen muuta* 'and primarily'	13
19	*ja Ranskan kanssa* 'with … and France', gen.	13
20	*että Suomen ja* 'that … of Finland and', gen.	12
	Number of clusters in total	484

clusters based on the Soviet Era translations, but mostly differs from it. Among the shared features should be mentioned the cluster at the top of the list—*Neuvostoliiton ja Suomen* 'the Soviet Union and Finland'—that names the main actors in the foreign policy of the neighbouring countries. As in the other corpus, Finland and Russia (or the Soviet Union) are represented in several clusters. Both of these acting states occur in four clusters with the element *ja* 'and'. The clusters with at least one of them together account for 10 positions in the list.

The repetition of the Soviet discourse is obvious, but even so there are some novelties in the post-Soviet list. In addition to the two neighbouring countries, there have appeared some extra states that co-operate with some other state. *France* occurs in three clusters, *England* in two, *Germany* in one. "The border discourse" has extended into a broader European space. Other newcomers in the list that should be mentioned are markers of the individual-oriented genre of history instead of the earlier official state-ruled terminology. The following items belong to this group of clusters: *ja sanoi että* 'and said that', *ja sen jälkeen* 'and after that', *ja niin edelleen* 'and so on', *ja ennen kaikkea* 'and first of all', *ja lisäsi että* 'and added that', *ja ennen muuta* 'and primarily'.

In addition to the differences in content of the keyword lists retrieved from the two corpora, the difference shows up in the frequencies of clusters. The total number of the clusters with the element *ja* 'and' in the top 20 list of the Soviet Era is 1696 (5.99 per 1000 words). The corresponding frequency in the post-Soviet corpus is 484 (2.50 per 1000 words). These numbers demonstrate that the translations from the Soviet Era are characterised by much more repetition.

2.3.4 Keywords in Narrative Structures

A more detailed keyword analysis can be conducted by applying larger contexts where individual keywords are examined. Our study uses as a method the so called *actant analysis*, which was developed by Greimas (1980) for studying narrative texts. This approach is based on Vladimir Propp's (1928) more complex model that was used for analysing Russian fairy tales. The choice of the method is justified by the character of history texts: they are textual realisations of narratives. According to Dray (1989: 111–112), narrativity is considered as one the focal features of historical discourse. Greimas and Courtés (1982: 5) define *actant* as something "that accomplishes or undergoes an act, independently of all other determinations". The relevance of the actant model for analysing manifestations of ideology is supported by the claim that "ideology can be defined as an actantial structure which actualises the values that it selects within axiological systems" (Greimas and Courtés 1982: 149).

Greimas's actant analysis is based on recognising the *actants*—the semantic roles of the actors in a narrative. The model consists of three opposed pairs of actants. The main concept is represented by the actants of *subject* and *object*. The subject is the acting hero in a narrative, the one who does something. The object is the goal towards which the action of the subject is oriented, something that is desired by the subject. Another pair of actants is formed by *helper* and *opponent*. The helper is understood as an actor giving assistance to the subject and trying to reach a common aim with the subject. The opponent, on the contrary, refers to the villain in a narrative. The opponent aims at preventing the subject from reaching the goal. Typical of every story is the double narrative trajectory: one for the hero and the other for the anti-subject (the villain). The third pair of actants in Greimas's model consists of *sender* and *receiver*. The receiver acts as an initiator of the action carried out by the subject. The receiver stands for the one who benefits from the subject's action (Greimas 1980: 201–207; Greimas and Courtés 1982: 371). The study of manifestations of the last mentioned pair is disregarded in the present analysis.

The actant analysis is conducted in the present study using a corpus-driven approach. The semantic roles are defined in the contexts where the keyword *Neuvostoliitto* 'the Soviet Union' occurs. The analysis is based on the use of concordances of the word *Neuvostoliitto* 'the Soviet Union', and it is examined in the two corpora representing different periods of time. The word occurrences are resorted according to the first word on the right (and on the left as well), which arranges the co-text elements in alphabetical order and helps to observe regularities in the neighbouring lexical items. Concordances provide raw material for analysing narrative structures. They can be used for demonstrating manifestations of *the semantic prosody*—"the consistent aura of meaning with which a form is imbued by its collocates" (Louw 1993: 157)—hidden in the context of the keyword in focus. According to Sinclair (1998: 20) and Stubbs (2001: 65), semantic prosody reveals attitudes of the language user towards something represented in the text.

The keyword *Neuvostoliitto* 'the Soviet Union' represents one of the main actors in the narrative of the political history of Finland. The narrative structure in the two sets of texts looks quite different. The history texts dating from the Soviet Era present the Soviet Union as a hero-subject, despite the fact that the texts discuss the history of Finland. An interesting detail is that in these texts the subject actant meets the helper actant. The Soviet Union co-operates with Finland and helps it to deal with foreign policy threats. The object of acting is to reach peace and harmony between the two neighbouring countries.

Example 1 ***Suomi*** *oli ensimmäinen teollisesti kehittynyt kapitalistinen maa, jossa ryhdyttiin* ***Neuvostoliiton*** *teknisen avun turvin rakentamaan atomivoimaloita*. (**Finland** was the first industrially developed capitalistic country that started to build nuclear power plants with the help of **the Soviet Union**.) (Bartenjev and Komissarov 1977)

The role of the Soviet Union as a helper is illustrated in example 1. The helper-subject, the leading socialistic state, assists a capitalistic country to get technically developed. Finland has acted very reasonably: it is the first country in the capitalistic block whose leaders have understood to start co-operation with the Soviet Union.

The lexical item verbalising the main subject of the Finnish history *Suomi* 'Finland' occurs, naturally, in the close environment of the keyword *Neuvostoliitto* 'the Soviet Union'. Although the latter has taken the place of the hero-subject and helper, Finland is still needed for constructing a story.

Example 2 Samaan aikaan kun ***Suomen*** *hallitus julisti uskollisuuttaan puolueettomuudelle,* ***Suomen*** *kenraalikunta kehitteli sotilaallista oppia, jonka lähtökohtana oli yksi ainoa mahdollisuus – sota* ***Neuvostoliittoa*** *vastaan*. (At the same time as the Government of **Finland** proclaimed its loyalty to neutrality, the generals of **Finland** were developing a military doctrine that was based on one single possibility—a war against **the Soviet Union**.) (Komissarov 1974)

Example 2 introduces the other semantic role of Finland in addition to the co-operating hero—the role of an opponent. A typical feature for a narrative is the earlier mentioned double trajectory. This means that there are two different narratives in a text at the same time. The writer takes a position on the side of the hero-subject who is trying to reach a certain object. He builds the text from the point of view of the hero. This story is accompanied by another narrative, the one of an opponent, also labelled as *anti-subject* or *opponent-subject* (Greimas and Courtés 1982: 371).

Actually, this example demonstrates the double narrative trajectory very clearly. The first item *Suomen*, referring to Finland, displays the role of the hero-subject. It represents the discourse of Finland's official foreign policy. The opponent-subject is verbalised by the second, identical lexical *Suomen* element with the meaning 'Finland'. The Soviet discourse points out the enemies of the friendship policy, in this case the Finnish generals.

The translations of texts written after the collapse of the Soviet Union show how the semantic roles in the narratives have changed. The keyword *Neuvostoliitto* 'the Soviet Union' appears as a verbalisation of the subject, but its status has changed. It is no longer the hero of the story and not a helper. The subject makes mistakes, it could be described as an anti-hero, maybe even a victim-hero. However, the Soviet Union is not described clearly as an opponent-subject either.

Example 3 ***Neuvostoliiton*** *hallitus teki pahan virheen, kun se vuoden 1940 joulukuussa ei asettunut tukemaan…* ***Suomen*** *liittoutumista Ruotsin ja Norjan kanssa maiden puolueettomuuden turvaamiseksi Saksan mahdollisilta loukkaamisyrityksiltä.* (The Government of **the Soviet Union** made a bad mistake when it did not in December 1940 take a position in supporting… **Finland's** association with Sweden and Norway for the sake of guaranteeing neutrality in these countries against the possible attempts of violation by Germany.) (Sinitsyn 1995)

Example 3 gives briefly an idea of the role of the Soviet Union as a subject in a narrative. The Soviet Union represents the old regime in the post-Soviet reality and acts in the role of an anti-hero. Its decisions are criticised and, on the contrary, Finland's efforts in reaching the goal of neutrality are highlighted. This discourse would have been impossible in the texts of the Soviet Era, where there is no criticism towards one's own government, and instead of mentioning Finland's negotiations on neutrality, the author of the text would have stressed Finland's contacts with Nazi Germany—acting in the role of the opponent.

The last example (4) in principle places the Soviet Union in the category of the opponent-subject. Nevertheless, the role is not defined by the writer of the story. The narrator transfers the responsibility from the Soviet Government to the public opinion that defined the role of the Soviet Union as an invader. Finland is introduced here as a victim-hero.

Example 4 Talvisodan ensi päivistä saakka koko maailman yleinen mielipide julisti ***Neuvostoliiton*** *hyökkääjäksi ja asettui* ***Suomen*** *puolelle.* (From the first days of the Winter War the public opinion of the whole world declared the Soviet Union as an invader and took sides with Finland.) (Sinitsyn 1995)

This example clearly shows that both acting states have taken a new role compared to the texts of the Soviet Era. The Russian authors of the newer books on the history of Finland have more freedom in choosing their perspective. That is why the translations also reflect a broader range of evaluations on Finnish history than the Soviet Era translations could do. The latter, instead, provided stereotype roles of Finland and the Soviet Union.

2.4 Conclusions and Discussion

The aim of the study was to introduce a corpus-based keyword analysis of translated history texts and combine it with a study of the narrative structure of these texts. The corpora, consisting of translations from two time periods with distinct features of the prevailing text production environment, showed differences in keywords. The corpora from the Soviet Era were characterised by frequent repetition of keywords referring to the two main actors in the narratives of the Finnish political history—Finland and the Soviet Union. In addition, these texts focus on verbalising the semantic field of positive relations between the main actors. Contents such as, ‘friendship’, ‘co-operation’, and ‘agreement’ are typical of the keyword list retrieved from the Soviet Era corpus.

The post-Soviet corpus of translations presents a keyword list with a range of grammatical words instead of reoccurring content words typical of the Soviet Era texts. The post-Soviet translations give more space for individual use of words, and they are distinguished from the texts in the other corpus by their references to the first person singular and to human actors in general. The texts from the Soviet Era are characterised by the action of states, armies, governments and other official issues.

The keywords were examined in larger contexts by using cluster analysis. The study was focused on the clusters that shared the only common keyword *ja* ‘and’ in the keyword lists of both translation corpora. The cluster analysis revealed the same kind of differences between the two corpora as the keyword lists. The Soviet Era clusters repeat the contents familiar from the keyword list. The coordinating conjunction *ja* ‘and’ links the verbalisations of positive relation between the actors into clusters, such as *Neuvostoliitto ja Suomi* ‘the Soviet Union and Finland’, *ystävyyden ja yhteistyön* ‘friendship and co-operation’ and *rauhan ja turvallisuuden* ‘peace and security.

The next part of the study, the actant analysis, made it possible to show how the keywords are linked together in a narrative structure. Combining the information taken from specific statistical data with qualitative considerations enables a link to be built between the two levels of ideology: the linguistic markers and the social context. The latter is needed for the interpretation of the markers.

The results of the actant analysis make it clear that there exists a difference between the narrative structures in the Soviet Era and post-Soviet Era corpora. The Soviet Era texts introduce a co-operating hero—*Finland and the Soviet Union*—reaching the goal of friendship. The semantic role of the Soviet Union is the subject and the helper at the same time. Finland, especially its right wing and capitalistic forces, fulfils another role as well—the role of the opponent. The post-Soviet texts illustrate a more complicated picture of actants, where the Soviet Union can appear as an anti-hero and Finland as a victim of the action of the Soviet Union.

The method of combining corpus statistics and qualitative analysis give new challenges for further research. The linkage between the statistical features of translated texts and the reception of translations appears especially interesting. Mäkisalo and Kemppanen (2010) carried out a reception test where the subjects were asked to rank Russian-Finnish translations of the political history of Finland and non-translated texts of the same genre in terms of the foreignness versus the domesticity of the texts. It was hypothesised that the concept of keyness could be used for operationalising the concepts of foreignisation and domestication. Despite the fact that in this particular study there was no significant correlation between the results of the corpus statistics and the reception test, it is clear that approaches that combine different methods provide new perspectives for corpus-based translation and linguistic studies.

References

Baker, Mona. 1993. Corpus linguistics and translation Studies. Implications and applications. In Text and Technology. In Honour of John Sinclair, eds. Mona Baker, Gill Francis, and ElenaTognini-Bonelli, 233–250. Philadelphia and Amsterdam: John Benjamins.

Baker, Mona. 1995. Corpora in translation studies: An overview and some suggestions. *Target* 7 (2): 223–243.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager, ed. Harold Somers, 175–187. Amsterdam: John Benjamins.

Baker, Mona. 1999. The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics* 4(2): 281–298.

Bartenjev, T., and J. Komissarov. 1977. *Kolmekymmentä vuotta hyvää naapuruutta (Thirty Years of Relations of Good Neighbourhood)*. Trans. U. Heino. Helsinki: Otava.

Biber, Douglas, and Susan Conrad. 1999. Lexical bundles in conversation and academic prose. In Out of Corpora. Studies in Honour of Stig Johansson, eds. Hilde Hasselgård, and Signe Oksefjell 181–190. Amsterdam: Rodopi.

Chesterman, Andrew. 2011. Translation universals. In Handbook of Translation Studies, ed. Yves Gambier, and Luc van Doorslaer, vol. 2, 175–179. Amsterdam and Philadelphia: John Benjamins.

Dray, William H. 1989. *On history and philosophers of history*. Leiden, New York, København and Köln: E.J. Brill.

Fairclough, Norman. 1995. *Critical discourse analysis: The critical study of language*. London: Longman.

Fairclough, Norman. 1997. Miten media puhuu? (Media Discourse). Trans. V. Blom, and K. Hazard. Tampere: Vastapaino.

Firth, J.R. 1968 [1957]. A synopsis of linguistic theory, 1930–55. In Selected Papers of J.R. Firth 1952–59, ed. Frank R. Palmer, 168–205 London: Longmans.

Firth, J.R. 1969 [1957]. The technique of semantics. Papers in Linguistics 1934–1951. Reprinted. London: Oxford University Press, 7–33.

Fowler, Roger, and Gunther Kress. 1979. Critical linguistics. In *Language and control*, ed. Roger Fowler, Robert Hodge, Gunther Kress, and Tony Trew, 185–213. London: Routledge and Kegan Paul.

Greimas, Algirdas J. 1980. Strukturaalista semantiikkaa (Sémantique structurale. Recherche de métode). Trans. E. Tarasti. Helsinki: Gaudeamus.

Greimas, Algirdas J., and Joseph Courtés. 1982. Semiotics and Language: an Analytical Dictionary (Sémiotique: Dictionnaire raisonné de la théorie du langage). Trans. L. Crist, et al. Bloomington: Indiana University Press.

Hatim, Basil. 1999. The cultural and the textual in the way translation studies has evolved. Paper presented at the Seminar held at the Centre of Language and Linguistics of the University of Salford, 24 March 1999.

Hatim, Basil, and Ian Mason. 1997. *The Translator as Communicator*. London and New York: Routledge.

Heikkinen, Vesa. 1999. Ideologinen merkitys. Kriittisen tekstintutkimuksen teoriassa ja käytännössä. (Ideological Meaning in the Theory and Practice of Critical Text Analysis). Helsinki: Suomalaisen kirjallisuuden seura.

Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Jantunen, Jarmo H. 2004. Synonymia ja käännössuomi. Korpusnäkökulma samanmerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. (Synonymity and Translated Finnish. A Corpus-based View of Contextuality of Synonymous Expressions and Lexical Features Specific to Translated Language). University of Joensuu Publications in the Humanities 35. Joensuu: University of Joensuu.

Kellner, Douglas. 1998. Mediakulttuuri. (Media Culture: Cultural Studies, Identity and Politics between the Modern and the Postmodern). Trans. R. Oittinen, et al. Tampere: Vastapaino.

Kemppanen, Hannu. 2004. Keywords and ideology in translated history texts: a corpus-based analysis. *Across Languages and Cultures* 5(1): 89–106.

Kemppanen, Hannu. 2005. Avainsanoja, ideologiaa ja lojaaliutta: venäjästä suomennettujen historiatekstien korpusanalyysi (Keywords, ideology and loyality: a corpus-based analysis of history texts translated from Russian into Finnish). In Käännössuomeksi. Tutkimuksia suomennosten kielestä, eds. Anna Mauranen and Jarmo H. Jantunen. Tampere Studies in Language, Translation and Culture. Sarja A, vol. 1, 185–208. Tampere: Tampere University Press.

Kemppanen, Hannu. 2006. Ideologija ili interferencija? Korpusnyj analiz russko-finskih perevodov i original'nyh finskih tekstov po političeskoj istorii Finljandii (Ideology or interference? A corpus-based analysis of Russian-Finnish translations and original Finnish texts on the political history of Finland). Universitetskoe perevodovedenie. Sedmye Fëdorovskie čtenija. Vypusk 7:221–232. Sankt-Peterburg: Peterburgskij gosudarstvennyi universitet. Filologičeskij fakul'tet.

Kemppanen, Hannu. 2008. Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi (Keywords and Ideology: A Corpus-based Analysis of Translated and Non-Translated History Texts). University of Joensuu Publications in the Humanities 51. Joensuu: University of Joensuu.

Kenny, Dorothy. 2000. Lexical hide-and-seek: Looking for creativity in a parallel corpus. In Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects, ed. Maeve Olohan, 93–104. Manchester, UK and Northamton MA: St. Jerome Publishing.

Knowles, Murray, and Kirsten Malmkjær. 1989. Translating ideology: Language, power and the world of the Tin Soldier. In *English Language Research Journal LR Journal 3: Language and Ideology*, ed. Murray Knowles, and Kirsten Malmkjær, 205–241. Birmingham: University of Birmingham.

Komissarov, Juri. 1974. *Suomi löytää linjansa. Neuvostoliittolaisen historiantutkijan näkemys Suomen tasavallan vaiheista (Finland Finds Its Course. The Point of View of A Soviet History Scholar on the Phases of the Finnish Republic)*. Helsinki: Otava.

Laviosa, Sara. 2000. TEC: A resource for studying what is "in" and "of" translational English. *Across Languages and Cultures* 1(2): 159–177.

Laviosa, Sara. 2002. Europe in the making in translational and-non translational English. In CTIS Occasional Papers, volume 2, ed. Keith Harvey, 85–95. Manchester: Centre of Translation and Intercultural Studies, UMIST.

Laviosa, Sara. 2003. Corpus-based translation studies: Where does it come from? Where is it going? A paper read at the conference Corpus-based Translation Studies Research and Applications Conference, University of South Africa, Pretoria, 21–25 July 2003.

Laviosa-Braithwaite, Sara. 1996. The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation. Department of Language Engineering, Volume I. A thesis submitted to the University of Manchester Institute of Science and technology for the degree of Doctor of Philosophy.

Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Text and Technology. In Honour of John Sinclair, eds. Mona Baker, Gill Francis and Elena Tognini-Bonelli, 157–176. Philadelphia and Amsterdam: John Benjamins.

Mäkisalo, Jukka, and Hannu Kemppanen. 2010. Frekvenssejä ja reseptiota: Kaksi näkökulmaa kotouttamiseen ja vieraannuttamiseen (Frequencies and reception: two perspectives on domestication and foreignisation). In MikaEL—The Publications of the Annual Finnish Symposium in Translation and Interpreting Studies, 4: 1–14. http://www.sktl.net/mikael/vol4.php. Accessed 25 June 2013.

Mauranen, Anna. 2000. Strange strings in translated language. A study on corpora. In Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects, ed. Maeve Olohan, 119–141. Manchester: St. Jerome.

Mauranen, Anna, and Pekka Kujamäki (eds.). 2004. *Translation universals—Do they exist?*. Amsterdam and Philadelphia: John Benjamins.

Nelson, Mike. 2000. A corpus-based study of business English and business English teaching materials. Unpublished thesis. University of Manchester: Manchester. http://users.utu.fi/micnel/thesis.html. Accessed 25 June 2013.

Olohan, Maeve. 2000. *Introducing corpora in translation studies*. London and New York: Routledge.

Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London and New York: Routledge.

Petrilli, Susan. 1992. Translation, semiotics and ideology. *TTR* 5(1): 233–264.

Propp, Vladimir. 1928. *Morfologija skazki (Morphology of the Folktale)*. Leningrad: Academia.

Puurtinen, Tiina. 1998. Syntax, readability and ideology in children's literature. *Meta, XLII* 4: 524–533.

Sinitsyn, Jelisei T. 1995. Vaiettu totuus: salaisen agentin todistajalausunto (The Untold Truth: A Witness Statement of a Secret Agent). Trans. E. Adrian. Helsinki: Otava.

Scott, Mike. 1998. WordSmith Tools Manual. Version 3.0. Oxford: Oxford University Press. http://www.lexically.net/wordsmith/version3/manual.pdf. Accessed 25 June 2013.

Scott, Mike. 2014. WordSmith Tools Manual. Version 6.0. Oxford: Oxford University Press.

Scott, Mike, and Christopher Tribble. 2006. *Textual patterns. Key words and corpus analysis in language education*. John Benjamins: Amsterdam and Philadelphia.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John. 1998. The lexical item. In Contrastive Lexical Semantics, ed. Edda Weigand. Amsterdam Studies in the Theory and History of Linguistic Science. Series IV, Current Issues in Linguistic Theory, 1–24. Amsterdam and Philadelphia: John Benjamins.

Sowarka, Bernhard H. 1987. Die Ideologiehaftigkeit kognitiver Strukturen in Texten. In *Textanalyse als Kognitionskritik? Möglichkeiten und Grenzen ideologiekritischer Inhaltsanalyse*, ed. Peter Vorderer, and Norbert Groeben, 46–136. Tübingen: Narr.

Stubbs, Michael. 1996. *Text and corpus analysis*. Computer-assisted Studies of Language and Culture. Oxford: Blackwell.

Stubbs, Michael. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Thompson, John B. 1994 [1990]. Ideology and modern culture: Critical and social theory in the era of mass communication. Reprinted. Cambridge: Polity Press.

Vološinov, V.N. 1972 [1930]. Marksizm i filosofija jazyka. Osnovnye problem sociologičeskogo metoda v nauke. (Marxism and the Philosophy of Language. Fundamental Problems of the Sociological Method in the Science of Language). Janua Linguarium. Series Anastatica, 5. 2nd edition. The Hague and Paris: Mouton.

Williams, Raymond. 1976. *Keywords: A vocabulary of culture and society*. London: Fontana/Croom Helm.

Williams, Ian A. 2009. A corpus-based study of Spanish translations of the verb 'report' in biomedical research articles. *Meta* 54(1): 146–160. http://id.erudit.org/iderudit/029798ar. Accessed 25 June 2013.

Zanettin, Federico. 2012. *Translation-driven corpora. Corpus resources for descriptive and applied translation studies*. Manchester, UK and Kinderhook (NY), USA.

Chapter 3
Keywords—A Tool for Translation Analysis

3.1 Introduction

This chapter gives a methodological overview of *keyword analysis*—a corpus-based method applied in translation studies for examining lexical features of comparable texts. It introduces the concept of keywords and characterizes different types of materials and research questions relevant for keyword analysis.

The methodological section will attempt to combine some of the results achieved in using the keyword method not only in translation studies but also in linguistics and literary studies. This section was inspired by observations about the use of reference literature. Scholars in translation studies, linguistics and literary studies use the same corpus tools, including the keyword function, but rarely refer to each other across the boundaries of their research fields. This is still evident today in research practices, although translation studies scholars and linguists have shown tendencies to explore common ground as well. Especially scholars in contrastive linguistics and translation studies have found shared research interests in corpus-based studies, as stated by Laviosa (2011: 21–22). As an example she gives the *UCCTS conferences* organized in Shanghai in 2007 and in Ormskirk in 2010, where researchers in contrastive linguistics and translation studies gathered together to exchange views and ideas. Translation studies can especially make use of linguistic oriented research when combining corpus methods with discourse analysis. That is the reason for introducing here also keyword research that has been done outside translation studies in the fields of corpus stylistics and corpus-based discourse analysis.

S. Laviosa et al., *Textual and Contextual Analysis in Empirical Translation Studies*, New Frontiers in Translation Studies,
DOI 10.1007/978-981-10-1969-2_3

This chapter will show different perspectives of the keyword method. The main focus is in applying this method in translation research, but the chapter also introduces such applications of keywords that can be used, in addition to research purposes, for practical aims as well. This approach is derives from the pedagogical aims of this book. Aquaintance with a research method that can also be applied in practical work will presumably motivate students in learning corpus tools.

The methodological part will thus discuss the following issues: the concept of keywords (Sect. 3.2),the implementation of keyword analysis (Sect. 3.3), corpus software (Sect. 3.4), different materials and different approaches (Sect. 3.5), large corpora and translated language (Sect. 3.6), style in the original and in the translated text (Sect. 3.7), corpus-based discourse analysis (Sect. 3.8), sorting the retrieved keywords (Sect. 3.9) and examples on analysing "the aboutness" of a text (Sect. 3.10).

3.2 Concept of Keywords

Corpus-based retrieval of keywords is a relatively new issue in translation studies. It has been used as a part of corpus analysis since the late 1990s and the early 2000s, when "the corpus linguistic turn" was taking place in translation research, as this era has been referred to by Laviosa (2003). The same time-period is characterized by rich development in the field of linguistic corpus studies as well, as Culpeper (2012: 81) notes in her reference to the history of the corpus-based keyword method.

Nevertheless, the concept of keywords has its roots in the pre-corpus era of linguistic sciences. As stated already in Chap. 2, the idea of cultural keywords was introduced in 1930s by J. R. Firth (1968 [1957], 1969 [1957]). He emphasised that words had to be analysed in their actual contexts (1969 [1957]: 10). Firth's groundbreaking works were followed by Williams's (1976) dictionary of keywords in British culture. In this study identification of keywords was based on intuitive knowledge about the keyness of English words. Intuitive considerations formed the basis for identification of keywords also in the first corpus-based studies on cultural keywords carried out by (Stubbs 1996, 2001).

Firth's view on keywords stresses the significance of certain words in a certain culture. The cultural aspect has been attached to keywords also later in the history of applying this concept. This fact is interesting from the point of view of a research theme presented in the second chapter of this book. Keywords are expected to reveal something crucial about the culture, society and ideology behind the node text or set of texts. Corpus linguistic studies have, however, shown that keywords can expose other features of texts as well. Section 3.9 in the present chapter will take a look at different kinds of keywords.

The Firthian idea of words that are "key" in a certain culture was developed further in corpus-based keyword analysis. However, the first generation of corpus analysts was still leaning on intuitive recognition of keywords. Rayson (2012: 179–180) mentions such scholars as Stubbs (1996), Wierzbicka (1997), Ooi (2000) and Teubert (2001) who intuitively selected keywords that they considered culturally important and, for this reason, relevant for a closer analysis. The use of the chosen words was examined in collocation analyses. Rayson (2012: 180) points out that despite the fact that key elements were retrieved without any objective discovery process, frequency information still played an important role in choosing words for analyses. The analysts checked if the keywords in focus were common words in the target corpus.

A new, objective approach to keywords was invented by Mike Scott who included the keyword function in his famous corpus program *WordSmith Tools*. Scott (2014) states in the manual of the program that "the term *key word*, though it is in common use, is not defined in linguistics. This program identifies key words on a mechanical basis by comparing patterns of frequency." According to Scott "a word is said to be "key" if: (a) it occurs in the text at least as many times as the user has specified as a Minimum Frequency, and (b) its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p value specified by the user." *Appropriate procedure* refers here to statistical computing that includes the chi-square test or the Log Likelihood test.

3.3 Keyword Analysis—A Corpus-Based Method

The use of the keyword method requires three basic elements: suitable computer software, appropriate research material and a relevant research question or questions to answer. Keyword analysis is based on the use of a software function provided by corpus programs, such as *WordSmith Tools* and *AntConc*. As in all corpus-based reseach, the material should be available in electronic form and readable by the computer program (usually saved as text files).

Keyword analysis has been used as a method in several fields of research. This section makes an effort to gather theoretical knowledge on using the keyword method primarily in translation studies, but relevant information has been collected outside this area as well. As a matter of fact, when one looks at the references in research literature on corpus methods, it is suprising to notice how isolated from each other linguists and translation researchers are in their work. Spreading knowledge across disciplines would be fruitful for finding new research ideas.[1]

[1]This fact is probably due to the development of translation studies as a clearly separate research field since the 1990s. The first generation of corpus scholars in translation studies referred a lot to linguistic corpus studies because of the lack of translational research in this field. The newer generation continues to refer to classics in linguistics, such as Sinclair and Biber, but seems to ignore a lot of relevant sources in present-day linguistic corpus research literature. In linguistic corpus studies for its part, it is typical that translational corpus research is totally missing in the list of reference literature.

The term *keyword analysis* refers, firstly, to generating a keyword list and making conclusions from the information given by the list (e.g., Xiao and McEnery 2005). Secondly, it has also been used for naming a method that can consist of several analyses applying different corpus-based tools, such as wordlists, collocations, word clusters and concordances), or combining corpus methods with qualitative analysis of text materials. Here a keyword list acts as a starting point for further analyses of the contexts where the keywords in focus occur. As an example Kemppanen's (2004, 2008) studies where corpus methods are combined with a narrative analysis of texts on political history could be mentioned. This method has been labelled in research literature also with the term *keywords analysis* (Olohan 2004, Baker 2006, 2010) or often simply just *keywords*.

It is an interesting fact that outside linguistic sciences the term *keyword analysis* is nowadays frequently used for referring to search term research which is used for commercial aims, for example in marketing. So, the concept of keyness plays an important role not only in linguistic research, but also in the production and reception of texts in various fields.

3.4 Corpus Tool

The first of the basic elements of a keyword analysis is the corpus software. In many cases the use of the same corpus program is the only linking element between keyword studies carried out in different research fields. Acquaintance with keyword analyses sharing the use of the same software, even if they had been done outside translation studies, can generate new research ideas.

The computer program should be able to generate word lists from comparable text materials and make a keyword list by comparing statistical features of the two word lists.

As Römer and Wulff (2010: 103) state, three of the most commonly used software packages for corpus analysis are *WordSmith Tools*, *MonoConc Pro*, and *AntConc*. All of these sets of programs run a keyword function, although in *MonoConc Pro* the corresponding function is labelled *corpus comparison*. *WordSmith Tools* is the most well-known and widely used of these sets of programs. They are applied in all of the fields that are relevant for us: translation studies, linguistics and literary studies. *AntCont* has recently been challenging *WordSmith* and *MonoConc* for one reason: it is a freeware, compared to the two commercial packages. As far as the keyword function is concerned, *WordSmith Tools* and *AntCont* are more advanced compared to *MonoConc Pro*. They are able to provide a ready-made keyword list on the basis of automatic calculations. *MonoConc*, in turn, shows the two target word lists with statistics in a window, where the analyst makes the comparison him- or herself.

It is typical that corpus studies scholars do not even mention any alternative programs for *WordSmith Tools* software, although this program is often given just as one example of existing programs. Descriptions of using *AntConc* are given by the author of the software Laurence Anthony in his two articles (Anthony 2005, 2006), by Römer and Wulff (2010) and by Maher et al. (2008). Wilkinson (2012) has tested this program as a translator's tool and found it very user-friendly compared to other freeware programs. As the only weakness of *AntConc* he mentions its tendency of being rather slow when using corpora of over 1000,000 words. So, in research work this might be considered as a shortcoming if you are analyzing universal or genre features of translations. When analyzing smaller corpora consisting, for example, of texts in a narrow specialized area or of individual texts, *AntConc* will be able to handle the material fast enough.

The choice of the appropriate software depends on the aims of the researcher. If the analyst is interested in more advanced study of keyness, where grammatical and semantic annotation is needed, it is advisable to use *WMatrix*. It accomplishes annotation automatically and generates a keyword list. Within *WordSmith Tools* you would have to do grammatical and semantic tagging separately before feeding the material into the keyword program function. Culpeper (2012) has applied *WMatrix* in her analysis of the character-talk of Shakespeare's *Romeo and Juliet*.

Xiao and McEnery (2005) proved in their study that *WordSmith Tools* is an especially handy instrument to be used in genre analysis. They compared Biber's multidimentional analysis (MDA) and Tribble's use of *WordSmith* in keyword analysis. As a conclusion Xiao and McEnery (2005: 77) state that *WordSmith Tools* is less demanding for the user compared to MDA, because it requires less expertise in data extraction and statistical analysis. MDA is a powerful tool for genre analysis, but it is complicated as well. The user-friendliness of the *WordSmith* software is one of the main arguments for choosing it. Especially, if you want to utilize keyword analysis for learning purposes as well as for research work, *WordSmith Tools* is a good alternative. Pearson (2000: 550) stresses the user-friendliness of the *WordSmith Tools* program in teaching work. She used this software for evaluating student translations with the help of specialized comparable corpora.

WordSmith Tools has been chosen for generating the keyword lists discussed in the methodological part and in the case study of the next chapter as well. The choice is justified by long personal acquaintance with this software, its user-friendliness and the fact that this well-known program package is usually accessible to corpus-oriented researchers and university teachers all over the world.

3.5 Different Materials and Different Approaches

As far as the research material is concerned, there are certain limitations to be taken into account, but even so the scope of appropriate comparable texts is endless. The present and the following sections will introduce different possibilities for using keyword analysis. The focus will be on the comparison of a single node text with a reference corpus consisting of a large number of texts. This focus has been chosen both for scholarly and pedagogical reasons. Examples on different analyses of single texts with the keyword method may give insights into new research themes in the field of corpus-based translation studies. Using just a single node text makes it easy to start working with new research themes and materials, especially if you already have a reference corpus that is suitable for several purposes. Designing and compiling a corpus with a great number of texts may not be feasible if you are working, for example, on your BA or MA thesis. The time schedules of the research process on these levels just do not allow for the time-consuming process of corpus compilation.

It is clear that choosing a single text may cause problems as well. You always need a good reason for research work. A one-text node corpus makes it easy to get aquainted with the keyword function as a method for both research work and practical examining of a text. This approach is a good example of *corpus-driven analysis* where you start from the text material without any theoretical assumptions and after generating a keyword list the study will continue on the basis of assumptions arising from the research material (cf. Tognini-Bonelli 2001: 84–87). Using a one-text node is an appropriate way of developing skills in mastering the keyword method as a translator's tool in practical work. In other words, acquaintance with this type of analysis strengthens connections between translation research and the practical work of translators. The focal role in this case will be given to source text analysis which can be applied for developing translation skills that help the translator to get a quick view of the key elements of each text and cope even with challenging tasks. An approach opposite to the corpus-driven one is the *corpus-based approach*. According to Tognini-Bonelli (2001: 65) "the term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study."

However the strict distinction between these two, seemingly contrasting approaches has been denied by several corpus scholars. McEnery et al. (2006: 8–11) state that a sharp distinction between corpus-based and corpus-driven analyses cannot be made because of the fuzziness that linguistic studies have shown in using these concepts. Mahlberg (2013: 13) and McIntyre (2007: 572–574) have criticized the principle, applied in corpus-driven research, of not using annotation in reseach material. They consider that especially in stylistic research there are a lot of features in a text that cannot be reached via computational means. They point out that a corpus-driven approach does not mean the same as a theory-free approach. The researcher cannot avoid using theoretical assumptions when analysing corpus material. Nonetheless, as Mahlberg (2013: 13) puts it, "what a corpus-driven way aims for, however, is to keep the assumptions minimal (cf. Mahlberg 2005)."

The choice of material is closely related to defining the research questions. It is important that texts which are taken for comparison on the one hand share certain features, but on the other hand are characterized by potentially differing features as well.

The main requirement for comparability of texts is that of mono-linguality of the material. The comparable materials for a single keyword analysis must consist of texts written in one language. In other words, a keyword list can be generated when you have monolingual material consisting of a target and a reference corpus. The idea of analyzing comparable monolingual materials in translation studies was invented and later further developed by Baker (1993, 1995, 1996). Formation of new research questions brought about a shift in translation studies from equivalence-oriented analysis to the target text oriented approach.

A keyword analysis can be applied to corpora with differing numbers of texts. In this respect there are different requirements for target and reference corpora. A target corpus may consist of just one text or a large number of texts. If the research question deals with the distinguishing features of an individual text, the target corpus may consist of just a single text. The keyword analysis is able to reveal how this certain text differs from the norm represented by the reference corpus. This type of keyword research can be used for exposing translation- or translator-specific features of the target text, e.g., Olohan's (2004: 160–167) study on Venuti's literary translations.

The choice of the text materials for the reference corpus is always a question that bothers reseachers involved in compiling a corpus. However, as Culpeper (2012: 85) puts it, "there is no magic formula for making this decision". One of the main issues is the size of the reference corpus. Baker (2010: 134) discusses two very different views on this topic presented by Berber Sardinha and Scott. Berber Sardinha (2004: 101–103) claims that the larger the reference corpus the better—you will find more keywords in analysis. He recommends that the reference corpus should be at least five times as large as the target corpus. A totally different argument has been offered by Scott (2009: 11) who found out in his comparison of different reference corpora with doctor–patient interactions that there is no 'bad' reference corpus. He concludes that keyword analysis is a robust tool which makes it possible to point out the focal elements in a text or set of texts, even when the reference corpus represents another genre than the target corpus.

Nevertheless, for a more detailed analysis of the material you need a corpus that has been specially designed for the research project in focus. Culpeper (2012: 86) stresses that "the closer the relationship between target and the reference corpus, the more likely the resultant reference corpus will reflect something specific to the target corpus." It should be added that the regularity in using the same reference corpus for comparing text materials plays an important role. A constantly used reference corpus helps to make differences between texts or sets of texts visible.

A comparison of retrieved keyword lists might also reveal possible shortcomings in the reference corpus. Some observations in a recent study by Kemppanen and

Mäkisalo (2016) could be mentioned as an example of such deficiencies. The study revealed that a general corpus consisting of newspaper articles is not the best reference corpus for comparing lexical features of translated and non-translated texts of academic non-fiction on political history. It was hypothesized that a keyword list could show the distance of each text from the frequency norm represented by the newspaper corpus. It was also assumed that translations of non-fiction would show more keywords and higher keyness values and appear more "foreign" compared to non-translated texts. The study revealed that keyword analysis could not make clear differences between translated and non-translated texts. Instead, the keyword lists generated both from translations and non-translations exposed the distance of the genre of non-fiction from the newspaper genre, regardless which variant of non-fiction each text represented—translation or non-translation.

3.6 Large Corpora and Translated Language

If the aim of a study is to describe a certain area of language use, the target corpus must be compiled of a great number of texts relevant to this research task. This kind of corpora has been very typical in translations studies. Following the tradition that was formed in the end of the 1990s and early 2000s, corpus-based translation research has mainly concentrated on analyzing lexical features of translated texts in comparison with non-translated texts.

Typical scholarly works using monolingual comparable materials are studies on *translation universals*. The first generation of translational corpus research focused on examining hypothesized universal features of translation, such as simplification, explicitation, normalization, standardization, levelling out, translation of unique items, untypical collocation and interference (see an overview in Zanettin 2012: 12–23). The first research groups focusing on the issue of universals were at the University of Manchester and at the University of Joensuu (Finland).

The role of keyword analysis in the works of the first corpus generation is quite marginal. Nevertheless, the method was used as a part of the studies by some of the scholars (e.g., Laviosa 2002 and Nevalainen 2004). Laviosa (2002) applies keyword analysis for introducing differences between the translated and non-translated texts in the Translational English Corpus. The focus of this study is in the use of vocabulary referring to the concept of 'European Union' (*European, Europe* and *Union*). Nevalainen (2004), for example, discusses differences between Finnish translations and non-translated texts in using colloqualisms. The keyword list showed clearly the difference between the two corpora. Compared to translations, non-translated texts were characterized by colloquialisms, such as dialectic variants of personal pronouns: *mä* or *mie* instead of literary variant *minä* 'I' and *sä* or *sie* instead of *sinä* 'you, sg. 2 person').

It was typical of the first generation studies that keywords were retrieved together with word lists, collocations, word clusters and concordances, and the method was applied for providing more evidence on universal features of translations. The comparable materials were chosen based on the idea of universals, which implies that the corpora consisted of translated and non-translated texts with various genres and translations from a great number of languages. An example of the great variety of a comparable corpus is provided by the structure of the *Corpus of Translated Finnish* (Mauranen 2000), which comprises texts of the following genres: fiction, children's fiction, popular fiction, academic prose and popular science. Although the total number of words in the two corpora is different—four million words of translated texts and six million words of non-translated texts—the corpora have approximitaly similar proportions of text materials representing each sub-corpus of different genres. The sub-corpora give an opportunity for testing hypotheses about regularities of translated language also on a lower level—on the level of a genre. The availability of sub-corpora representing translations from different source languages makes it possible to test hypotheses on the influence of the source language.

Despite the fact that corpus-based translation studies have later moved towards many new areas of study, most of the research in this field has concerned translation universals (Zanettin 2012: 23). Recent corpus studies have, however, already questioned the existence of universals. De Sutter et al. (2012: 138) point out two main reasons for the dissatisfaction of translation studies scholars with the Bakerian approach (e.g., House 2008, Becher 2010 and Bernardini and Ferraresi 2011). First, many of the classical hypotheses about certain universal features are "unmotivated, unparsimonious and vaguely formulated" (Becher 2010: 1), a fact that makes it difficult to interpret the results of corpus analyses. Secondly, corpus-based translation studies have largely ignored the possible influence of the source language and the genre variation. These claims against universals bring them into the focus of interest again. Questions about the influence of the source language and genre variation offer new themes for keyword analyses.

Actually, it is no wonder that the first generation of corpus researchers in translation studies were oriented towards universals and "translated language". The ideas on using corpora in translation studies were derived from the works of linguists who had started corpus-based research much earlier than translation studies scholars. At that time translation research had only a short history, so it was logical to study translations by using tools and ideas taken from linguistics. In addition, translation researchers themselves still had a background for the most part in linguistics rather than in translation studies.

3.7 Style in the Original Text and in the Translation

Another field for using the keyword method—in addition to the study of translation universals—is the study of style. This area has proved to be a great challenge for translation research. As Baker (2000: 243–244) writes, translation studies has inherited from linguistic and literary studies the association of style with 'original' writing. This has led to a late start for studies in translator's style. The idea that a translator could have a style of his or her own had to be discovered by translation theorists before further studies in translational stylistics were possible.

Kenny's (2001) analysis of literary texts by comparing German originals and their English translations from the point of view translators' creativity should be mentioned as a ground breaking study in examining translator's style. This study is also relevant for introducing the keyword method. Kenny generated separate keyword lists from both originals and translations by using two separate reference corpora—one for the German originals and one for their English translations. The keyword program function itself needs monolingual material, but it is, of course, possible to conduct separate keyword analyses by using materials in two or more languages and then compare the results of these studies.

Another example of a translation-oriented style analysis where the researcher applies the keyword method is introduced by Zanettin (2012: 120–121). He compares two translations of the same source text—Italian translations of Salman Rusdie's short story *Chekhov and Zulu*. One of the translations has been done by an anonymous translator and published in the online journal *Liberal*. The other one has been made by Vincenzo Mantovani and has been published in a collection of short stories. The keywords retrieved from the two translations suggest that Mantovani's text is more colloquial in style than the other one. For a more detailed study of hypothesized stylistic differences between the two translations, Zanettin recommends further investigations by looking at parallel concordances in both translations. Zanettin clearly points out the role of the keyword method as a starting point for a set of different corpus-based (quantitative) or qualitative analyses.

More new ideas for keyword analyses concerning the style of a translator can be obtained, for example, from works of the following translation studies scholars who have carried out experiments for quantifying 'translator style' by using corpus methods: Mikhailov and Villikka (2001), Munday (2008), Winters (2009) and Saldanha (2011). Especially Saldanha's study is a good help when discussing methodological difficulties in defining 'translator style'. She has analyzed different definitions of writers' and translators' styles and suggests a new definition for 'translator style'. It is "a 'way of translation' which: (1) is felt to be recognizable across a range of translations by the same translator, (2) distinguishes the translator's work from that of others, (3) constitutes a coherent pattern of choice, (4) is 'motivated', in the sense that it has a discernable function or functions, and (5) cannot be explained purely with reference to the author or source-text style, or as a result of linguistic constraints" (Saldanha 2011: 31).

In addition to research in translation studies, a source for innovative approaches to translational style research can be found in corpus stylistics. This field of study is represented by such scholars as Wynne (2006), Culpeper (2012) and Mahlberg (2013). This area is promising for an open-minded translation researcher who is ready to receive new ideas from a field that is relevant for analysing translations and their originals, but almost totally ignores the existence of translation studies in the common ground of knowledge. This claim is based on the lists of reference literature in the recent publications on corpus stylistics.

The following comment by Mahlberg (2013: 23) is, however, an encouraging exception: "One area, where aspects of stylistic effects become visible, is in literary translation. In the field of translation studies, source texts are compared with the help of parallel corpora. Methods from the field of corpora and translation studies may also prove useful for stylistic analysis. For instance, we may want to compare the results of a corpus stylistic study of an original with the results of a similar corpus stylistic study for a translation or different translations". Mahlberg has found common areas of interests in different research fields and carried out studies that break the boundaries between these fields. She has used corpus-methods, including keyword analysis, not just in corpus stylistics, but in corpus-based discourse analysis and translation studies as well (e.g., Mahlberg 2007a, b, 2009).

Ji (2010: 80–81) goes even further and argues for *corpus-based translation stylistics*—interdisciplinary study of translation style by using methodology and obtaining insights across different, but related research fields, such as computational stylometry (Hoover 2003), quantitative sociolinguistics (Biber and Finnegan 1994; Eckert and Rickford 2001) and cognitive stylistics (Semino and Culpeper 2002; McIntyre 2007). Ji finds it important that translation studies use research results from current literary studies, where applying computational techniques is a growing trend.

3.8 Corpus-Based Discourse Analysis

Corpus-based keyword studies have found a promising area for further studies in discourse analysis. It has to be mentioned, however, that corpus-based discourse analysis has not attracted particular interest among translation studies scholars. Gabrielatos's (2014) cross-disciplinary bibliography on this theme shows this very clearly. The list contains only a few translation-oriented studies in corpus-based discourse analyses, such as Romagnuolo (2009) and a new, still unpublished release by Zhang and Munday (forthcoming) entitled *Discourse Analysis and Translation*.

Discourse studies have close connections with other areas of corpus studies, so it is only logical that research innovations in translation studies can be expected from multidisciplinary studies combining theoretical knowledge in translation research with other fields. In some of the discourse-oriented studies, researchers take a clear stance towards the phenomena represented in the texts that are put under keyword analysis. These studies have been described by some scholars as *corpus-based critical discourse analysis* (Mulderrig 2011a, b).

Corpus-based discourse analysis has shared research interests with style researchers. According to contemporary stylistics, style is often verbalized in items which belong to a closed class of words, such as pronouns, conjunctions, prepositions and other function words. This group of key elements, the stylistic keywords (Mahlberg 2013; Scott 2014), is considered relevant from the point of view of discourse analysis as well. Baker (2006: 127) states that "[a]s the style of the text may play some role in the discourses within it, it is recommended that such high frequency words are not discarded at this point". He points out that they should not be removed outside the keyword analysis by creating 'stop lists' of grammatical words when using keywords software.

Applying the keyword method to discourse analysis brings us to multidisciplinary research fields. From the point of view of translation studies this approach is relevant when you are involved with translations produced under specific social, cultural or ideological conditions. When you use a methodology that goes beyond the linguistic appearance of texts, it is important to take into account Baker's (2004) warning regarding a lexical-only approach. He recommends, instead, a method that will combine both qualitative and qualitative analyses. This idea of a multi-method approach has been referred to as *triangulation* (Newby 1977: 123; McNeill 1990: 23). Baker (2006: 15–17) understands triangulation as being the combination of different corpus methods in one study, including e.g., keyword analysis alongside the generation of collocates, word clusters and concordances. A further step of triangulation requires the use qualitative methods together with corpus analyses. Discourse analysis, especially critical discourse analysis, is a good example of triangulation. Discourse studies have shared ground with sociology, ideology research and narrative studies.

3.9 Sorting Keywords

As stated earlier, there are two contrasting approaches for analysing a keyword list: corpus-based and corpus-driven analysis. A corpus researcher should be aquainted with both methods, because it is hard to draw an exact line between the two approaches, and, actually, it is not even needed. On one hand, in corpus-driven analysis it is impossible to examine corpora without any assumptions and background knowledge about the phenonmenon in focus and earlier studies in the field. On the other hand, when applying a corpus-based method you cannot avoid the fact that the analysis starts to push the study into a new direction, which may lead to building new hypotheses. As a conclusion, it could be stated that keywords can be sorted and categorized in different ways depending on the approach chosen by the analyst.

Words given by a keyword list have been divided into different functional groups according to the research interests of the scholars. In addition, categorizations reflect the theoretical background of each study. Mahlberg (2013: 67) has found that it is typical in keyword analyses to divide the keywords into three categories: (1) proper names, (2) words that indicate what the texts are about, and (3) words that relate to features of style. Scott (2000; Scott and Tribble 2006) has earlier suggested that keywords could be divided into two, or possibly into three groups. His own division of keywords consists of three groups that correspond to the categories mentioned by Mahlberg: proper nouns, aboutness keywords (content words) and stylistic keywords.

The category of proper nouns is usually skipped quickly in the analysis of keywords (e.g., Scott 2014; Baker 2006: 128). It is considered as an obligatory group of words that pops up in a list but does not play any important role in discourse production. Lexical items in the category labelled as ‘aboutness keywords’ is considered as more relevant to the analyst. As Baker (2006: 127) puts it, these words “tend to be lexical words: nouns, verbs, adjectives, adverbs and generally those which are more interesting to analyse.” It is clear that lexical words are focal elements in analysing the contents of the texts. The third group, the stylistic keywords, consists typically of grammatical words, such as pronouns, conjunctions, prepositions and other function words. The occurrence of these keywords reveals how the producer of the text combines the content elements together with these stylistic markers.

Nevertheless, sorting keywords on the basis of word classes does not provide the whole picture. As Baker (2010: 26) states, among the lexical words nouns and verbs are most useful when an analyst of a keyword list is making conclusions on the topic of the node text. Superlative adjectives, instead, emphasize that lexical words can be used not only for indicating the topic in a text, but also for emphasizing features of the described phenomenon.

Culpeper (2012: 90) has introduced another grouping of keywords—again into three, but in this case based on Halliday’s (1973, 1978, 1994) functional grammar. Following Hallidayan functions, she divides keywords into the following categories: ideational keywords, textual keywords and interpersonal keywords. Culpeper uses this categorization for analyzing language use of different characters in *Romeo and Juliet*. Ideational keywords characterize Romeo, textual keywords Mercutio and interpersonal keywords the Nurse. Culpeper stresses that “these are not discrete categories; they are designed to capture a functional emphasis.”

In addition to the aforementioned groupings, keywords have been divided into different categories on the basis of discourse analysis. Mahlberg (2009: 103–126), who has studied the use of corpora for learning about the language and discourse, divides keywords into the three following types according to their function in discourse represented by corpora of newspaper articles: (1) *cultural keywords*, (2) *textual keywords* and (3) *computer keywords*. This division reflects the readers’ ability vs. inability to intuitively define the key elements in a discourse. Mahlberg (2013) later applied Halliday’s grammatical functions for categorization of keywords in the same way as Culpeper (2012).

The first type of words labelled as *cultural keywords* refers to words that can be recognized intuitively as key elements of texts written in a certain culture. Items of this category were identified in keyword studies of the pre-corpus era. Keywords of the British culture were listed intuitively by Williams (1976) in his book entitled as *Keywords*. Williams (1976: 13) uses the term *keywords* with two different meanings. Firstly, “they are significant, indicative words in certain activities and their interpretation”, and, secondly, “they are significant, indicative words in certain forms of thought”. Williams’s study on keywords is highly relevant to the research questions in the present study because it focuses on practices and institutions related to the concepts of *culture* and *society*, which are pivotal issues in ideology research as well.

It is obvious that a great deal of cultural keywords can be intuitively recognized by a reader who is a specialist in analyzing cultural phenomena. Williams's (1976) vocabulary is a good example of this type of approach. Still, a corpus-based identification of culture-specific lexical elements gives the researcher, or in Mahlberg's case the learner, new possibilities in finding keywords describing the culture in focus. Using corpora makes it easy to analyze a huge amount of texts and get possible evidence for intuitive assumptions on the keyness of the designated elements. The idea of using a corpus method in defining cultural keywords was given by Michael Stubbs (1996).

It is clear that determining the set of words belonging to cultural keywords is crucial for translation studies with an orientation to critical discourse analysis (for example, the study presented in the next chapter), but there are still other key elements that play a role in discourse production. Mahlberg (2009: 109) distinguishes another type of keywords that he calls *textual keywords*. These words get there meaning in context and cannot be easily recognized by readers. As an example Mahlberg gives the use of the noun *move* referring to an action or political decision (for a more detailed explanation, see Halliday and Hasan 1976, Francis 1994, Partington 1998).

The third type of keywords categorized by Mahlberg (2009: 111) is *computer keywords*. This type refers to keywords retrieved by corpus software (e.g., *WordSmith Tools* or *AntConc*). The keyness of words is defined here purely on statistical grounds. For example, the *WordSmith Tools* software, which is the most used program in corpus analysis, calculates the keyness of words in the following way: the program computes the frequency of a word in the corpus under investigation, the number of running words in this corpus, the frequency of the same word in the reference corpus, and the number of running words in the reference corpus and then cross-tabulates this data. Statistical tests include the classic chi-square test of significance with Yates correction for a 2×2 table and Ted Dunning's Log Likelihood test, which gives a better estimate of keyness, especially when contrasting long texts or a whole genre against your reference corpus (Scott 2014).

3.10 Examples on Analyzing the Aboutness of a Text

This section introduces three examples of keyword analyses made on the basis of comparing a single node text with a reference corpus. The three node texts are academic articles that have been published in Internet journals (Bystrovskij 2003; Ahmednabiev 2007; Mar'janenko 2008). The numbers of words in the articles are 2 260, 1 465 and 2 120. The illustrations (Figs. 3.1, 3.2 and 3.3) should highlight the relevant features, though they show only the top of each keyword list. This is for practical reasons: (1) there would not be enough space for introducing a large number of keywords, (2) it is assumed that the very top of the keyword list will show more relevant facts concerning the aboutness of each text than the rest of the list.

saring.kws

File Edit View Compute Settings Windows Help

	Key word	Freq.	%	Freq.	RC. %	Keyness	P	Lemmas	Set
1	ПАЦИЕНТА	21	0,93	37		265,49	0,0000000000		
2	СЕМЕЙНОЙ	19	0,84	26		247,60	0,0000000000		
3	СЕМЕЙНАЯ	14	0,62	19		182,58	0,0000000000		
4	ВРАЧ	14	0,62	87		146,32	0,0000000000		
5	МЕДСЕСТРА	10	0,44	5		143,43	0,0000000000		
6	МЕДИЦИНЕ	12	0,53	35		141,66	0,0000000000		
7	СЕМЕЙНЫЙ	10	0,44	32		116,44	0,0000000000		
8	ВРАЧА	11	0,49	66		115,66	0,0000000000		
9	МЕДИЦИНСКОЙ	12	0,53	125		113,75	0,0000000000		
10	МЕДИЦИНА	9	0,40	30		104,15	0,0000000000		
11	СЕМЕЙНОГО	9	0,40	31		103,63	0,0000000000		
12	СЕМЬИ	15	0,66	622		102,08	0,0000000000		
13	МЕДИЦИНЫ	10	0,44	92		97,14	0,0000000000		
14	МЕДИЦИНСКОГО	10	0,44	94		96,74	0,0000000000		
15	ПОМОЩИ	15	0,66	1 187	0,02	83,19	0,0000000000		
16	НЕПРЕРЫВНОСТЬ	6	0,27	9		77,32	0,0000000000		
17	СПЕЦИАЛИСТ	8	0,35	115		70,89	0,0000000000		
18	ЗАБОЛЕВАНИЙ	7	0,31	58		69,38	0,0000000000		
19	ПАЦИЕНТ	6	0,27	21		68,91	0,0000000000		
20	ОБСЛУЖИВАНИЯ	8	0,35	148		66,99	0,0000000000		

Fig. 3.1 Keyword list based on Text 1 (Bystrovskij 2007)

t1ng.kws

File Edit View Compute Settings Windows Help

	Key word	Freq.	%	C. Freq.	RC. %	Keyness	P	Lemi
1	ЭКОЛОГИЧЕСКИХ	15	1,02	62		180,96	0,000000000	
2	ЭКОЛОГИЧЕСКОГО	12	0,82	30		155,23	0,000000000	
3	ПРИРОДЫ	14	0,96	164		141,74	0,000000000	
4	СТУДЕНТОВ	13	0,89	224		121,94	0,000000000	
5	ЭКОЛОГИЧЕСКОЙ	11	0,75	100		116,65	0,000000000	
6	ПЕДАГОГИЧЕСКОГО	6	0,41	8		83,59	0,000000000	
7	ПРИРОДЕ	9	0,61	169		82,89	0,000000000	
8	ДИСЦИПЛИН	6	0,41	12		79,80	0,000000000	
9	ЭКОЛОГИЧЕСКОЕ	5	0,34	7		69,29	0,000000000	
10	КУЛЬТУРЫ	11	0,75	1 527	0,02	58,32	0,000000000	
11	ЭКОЛОГИЧЕСКАЯ	5	0,34	29		57,21	0,000000000	
12	ОБРАЗОВАНИЯ	9	0,61	808	0,01	55,34	0,000000000	
13	ВОСПИТАНИЯ	5	0,34	57		50,85	0,000000000	
14	ЧЕЛОВЕКА	11	0,75	2 405	0,03	48,69	0,000000000	
15	ШКОЛЬНИКОВ	5	0,34	77		47,96	0,000000000	
16	ЭКОЛОГИИ	5	0,34	90		46,45	0,000000000	
17	ДЕЯТЕЛЬНОСТИ	10	0,68	2 381	0,03	42,63	0,000000000	
18	ПРОБЛЕМ	9	0,61	1 730	0,02	42,05	0,000000000	
19	ВУЗА	4	0,27	42		41,31	0,000000000	
20	ЭКОЛОГИЧЕСКИЕ	4	0,27	42		41,31	0,000000000	

Fig. 3.2 Keyword list based on Text 2 (Ahmednabiev 2007)

t3ng.kws

File Edit View Compute Settings Windows Help

	Key word	Freq.	%	. Freq.	RC. %	Keyness	P	Lemmas	Set
1	СПК	9	0,42	0		147,42	0,0000000000		
2	БАРИН	11	0,52	22		138,19	0,0000000000		
3	ПОМЕСТЬЕ	7	0,33	13		88,76	0,0000000000		
4	МЕНЕДЖЕР	6	0,28	35		64,15	0,0000000000		
5	ХУТОРА	4	0,19	3		55,95	0,0000000000		
6	БАРЫНЯ	4	0,19	8		50,24	0,0000000000		
7	НЕПРОДУКТИВНАЯ	3	0,14	1		44,63	0,0000000000		
8	ПРОДУКТИВНАЯ	3	0,14	2		42,40	0,0000000000		
9	ДЕВКИ	3	0,14	4		39,57	0,0000000000		
10	ТОВАРНОЕ	3	0,14	5		38,55	0,0000000000		
11	ПОМЕСТЬЯ	3	0,14	6		37,68	0,0000000000		
12	ОЧЕНЬ	13	0,61	5 660	0,07	32,08	0,0000000119		
13	ФАЗЫ	3	0,14	22		30,80	0,0000000257		
14	НАШЕ	6	0,28	640		30,53	0,0000000299		
15	ОРГАНИЗАЦИИ	8	0,38	2 160	0,03	26,63	0,0000002436		
16	ТРУДУ	3	0,14	51		25,99	0,0000003407		
17	ЭТО	25	1,18	27 675	0,36	24,43	0,0000007666		

Fig. 3.3 Keyword list based on Text 3 (Mar'janenko 2008)

It can be inferred from the bibliographical data of the articles that they represent three different fields of knowledge. More precise definitions of the aboutness of the texts will be left for corpus-driven keyword analysis. In addition to research purposes, the texts have been used as source texts during translation courses, so they fulfill a pedagogical function as well. It can be expected that the texts contain typical problems of LSP translation such as the presence of terminology related to each specialized field and of tacit knowledge packed between the lines by the authors of the tetxs. When training students in carrying out keyword analysis it is convenient to use materials that are applied also for some other purposes, not just for corpus training. Using the keyword tool in source text analysis when accomplishing a translation task is a good way to combine methodological training and practical translation work.

The reference corpus for all the three analyses consists of a 1 year volume of the Russian newspaper *Nezavisimaja gazeta* (1999) in electronic form. The corpus contains 245 newspaper articles; the total number of words is 7 million. The corpus is supposed to function as a frequency norm of general language use that each of the texts under investigation will be compared to. The statistically relevant differences between the frequencies in the word lists of the target corpora and reference corpora will be presented in the form of keyword lists.

A keyword list helps us to quickly to find out what a text is all about. As mentioned earlier, corpus scholars from different fields of language studies use the term *aboutness*—originally created by Philips (1989)—for this text feature (Scott 2000; Baker 2006; Culpeper 2012). This concept is useful both in translation research and in the practical work of translators. As Scott (2000: 110) puts it, keywords are one way of foregrounding information. High frequency of a word emphasizes the prominence of the concept labelled by this word in a certain text. Knowledge of the key elements of a text provides a good starting point for further analyses. Figure 3.1 provides an example of a keyword list generated from a text that was used as a material on a translation course. *Word Smith Tools* is an appropriate program for analysing the source text. Using the keyword function helps the translator to get a quick general view on the lexical features of the text.

The concept of aboutness is clearly brought out when working with corpora containing just one text. The keyword list introduced in Fig. 3.1 shows clearly two focal groups of content words that can be categorized by semantic criteria.[2] One group of words label medical concepts: *pacienta* (word #1), *pacient* (#19) 'patient', *vrač* (#4), *vrača* (#8) 'doctor', *medsestra* (#5) 'nurse', *medicine* (#6), *medicina* (#9), *mediciny* (#13) 'medicine', *medicinskoj* (#9), *medicinskogo* (#14) 'medical', *pomošči* (#15) 'help, care', *zabolevanij* (#18) 'disease'.

Five other words in the list form another semantic group: *semejnoj* (#2), *semejnaja* (#3), *semejnyj* (#7), *semejnogo* (#11) and *sem*'*i* (#12). The first four of them are different word forms of an adjective with a meaning 'family, with the family,' and they are used as the first part of compounds with a noun as the second part. The last one *sem*'*i* is a word form of a noun with a meaning 'family'. The word groups presented above reveal that the text is about medicine and, more precisely, about family medicine, a field with specific treatments emphasizing health care in the family context.

In addition to keywords referring to concepts of medicine and family, the list contains the word *nepreryvnost*' (#16)'continuity'. This keyword and the concept behind it cannot, at least directly, be associated with medicine. Anyone who analyzes the list is likely to wonder why this word is key. A closer acquaintance with the text reveals that the concept of continuity is crucial in family medicine. The case of *nepreryvnost*' 'continuity' is a good example of surprising keywords. Such items are problematic in terminological work. They cannot be easily inferred as focal elements of a special field if you are compiling a glossary without a keywords tool. They are often left outside the list of terms because of their common use in non-specialized texts. These words are not recognized as terms in a certain field.

It could be expected that such oversights when compiling a glossary would occur especially in the work of a compiler who maybe is a professional translator, but with limited experience in the specialized field. It is understandable that not even consultations with a specialist in the field would reveal lexical elements of this category. He or she would not probably consider them as terms of their field either. However, the language conventions in the field may require the use of a certain collocate from "general language use" as a part of a compound word. This causes problems especially for a language user who is working with a foreign language.

[2]The research material is non-lemmatized. That is why the word *word* is used here in the sense 'word form'. Russian nouns and adjectives take different case and gender forms (both singular and plural), the fact that is reflected in the composition of each keyword list. To help those readers, who do not know Russian, the keywords are provided with English translations. They are given for each lemma, not for different word forms.

He or she may choose an untypical collocate from the range of synonymous words that could function as possible equivalents of a word in the source language. Words in this category are clearly going through a process of shifting from general language use to specialized usage. Keyword analysis provides a handy instrument for recognizing term-like words that are useful in communication in specialized language.

The keyword list based on Text 2 introduced in Fig. 3.2 brings out, as the previous list did, what the text is all about by highlighting its essential terminology. This list shows again two semantic groups of keywords. The first one comprises words connected with the concept of ecology: *èkologičeskih* (#1), *èkologičeskogo* (#2), *èkologičeskoj* (#5), *èkologičeskoe* (#9), *èkologičeskaja* (#11), *èkologičeskie* (#20) 'ecological', *èkologii* (#16) 'ecology', *prirody* (#3), *prirode* (#7) 'nature'. The second group contains words referring to education: *vospitanija* (#13) 'education, upbringing', *obrazovanija* (#9) 'education, schooling', *vuza* (#19) 'university, college, polytechnic', *studentov* (#3) 'student', *škol'nikov* (#15) 'schoolchild', *disciplin* (#8) 'discipline', *pedagogičeskogo* (#6) 'pedagogical'. The keyword list clearly reflects the main contents of the text dealing with environmental education which should be implemented in all levels of schooling, beginning from the first school classes and continuing until the highest levels of academic education.

All the keywords in the top 20 list of Text 2 are content words clearly showing the aboutness of the text. Most of the keywords are terminological lexical items. Just three of the words could be characterized as items from general purpose language *kul'tury* (#10) 'culture', *problem* (#18) 'problem' and *dejatel'nosti* (#17) 'activity'.

The keyness of the words *èkologičeskij* and *èkologija*, highlighted by the corpus tool, informs the translator about the lexical elements that are central to the source text, and that require either knowledge of terminology in the special field or, if this is lacking, ability to find appropriate equivalents when rendering these words in each context. The keyword list draws the translator's attention to problematic items in the text. If a word has high keyness in the text, the translator should be very careful when choosing the appropriate equivalent for this word. An unsuccessful translation of a key element will probably cause more harm to the target text than an unsuccessful translation of some other word that is not a central one.

Translating the adjective *èkologičeskij* 'ecological' and the noun *èkologija* 'ecology' into Finnish is not as simple as it looks at first glance. In many cases the equivalents suggested by the bilingual Russian-Finnish dictionary (Kuusinen et al. 2007) *ekologinen* 'ecogical' and *ekologia* 'ecology' do not function as suitable options at all. Especially the adjective *èkologičeskij* 'ecological' may appear problematic in translation. It is typical for Russian language to build compounds by using an adjective and a noun (e.g., *èkologičeskoe vospitanie,* literally 'ecological education'). In Finnish such contructions are possible, but a more common way of building compounds is to put two nouns one after another and write the two parts of the compound together. Thus, in the case of *èkologičeskoe vospitanie* a conventional way to transfer this term into Finnish would be *ympäristökasvatus*, literally 'environment education'.

This text was translated after the keyword analysis session during a translation course. The translation process showed that key compounds with the noun *èkologičeskij* as the first part of each compound were challenging from the point of view of finding appropriate equivalents. These compounds required numerous Internet searches for creating conventional Finnish expressions in the field of environmental education.

The fact that a word appears as key reflects the importance of this item from the point of view of translation. The concepts labelled by these words are expected to require more detailed background work and more work in finding suitable equivalents as well.

The keyword list generated from Text 3 (Fig. 3.3) presents information about the source text differing from that in the previous lists. Some of the words in the list clearly belong to the field of economics: *SPK* (#1) 'socialist production collective', *neproduktivnaja* (#7) 'unproductive', *produktivnaja* (#8) 'productive', *tovarnoe* (#10) 'goods', *trudu* (#16) 'work, labor', *organizacii* (15) 'organization, organizing', *menedžer* (#4) 'manager'. On the basis of these keywords it is apparent that this text deals with development of the present-day Russian economy. However, some of the words in the list do not really fit into the overall picture of modern economics. These words form another semantic group consisting of the following items: *pomest'e* (#3), *pomest'ja* (11) 'estate', *hutora* (#5) 'solitary estate', *barin* (#2) 'estate owner', *barynja* (#6) 'lady of the manor', *devki* (9) 'maid'. All of these words share a common semantic feature: they label items belonging to a country estate of the Czarist era. Business operators in the modern Russian economy are compared with individuals running a pre-revolutionary Russian country estate. The list of keywords reveals in this case a metaphor hidden in the text. The translator or the translation studies scholar working on the source text analysis gets useful information on the overall structure of the text. The translator will pay attention to this specific structure and make a strategic decision whether or not to preserve the metaphor with its strongly culture-bound elements. The researcher, in turn, will get a list of key lexical elements in the source text. The study can be continued with an analysis of a translation of the source text. Has the metaphor structure been preserved? If not, how has it been changed in the translation process?

3.11 Concluding Remarks

This chapter aimed to introduce central theoretical and methodological issues about the concept of keywords and keyword analysis. It discussed the use of the keyword method not only in translation studies, but in other relevant research areas as well including corpus stylistics and corpus-based discourse analysis. The purpose of this multi-disciplinary study was to give new ideas for keyword research. The theoretical overview of the concept of keywords revealed that different fields of linguistic research have a common ground where keyword analysis can be developed further. It seems clear that translation studies scholars could adopt relevant knowledge and methodology from other fields of corpus-based language studies. Nevertheless, there are some problematic points in combining results of research areas that are seemingly close to each other. One of the main problems is the lack of common terminology. There are clearly overlapping elements in categorizations of keywords by scholars from different fields. For example, keywords represented under the notion of stylistic markers have a lot in common with those defined as discourse markers.

One area where keyword analysis has recently been applied, but was not discussed here, is the one of learner language studies. This research area has features that are relevant from the point of view of translation studies as well, and could for this reason be considered as a promising field in keyword analysis. Learning a second or a foreign language has a lot in common with translating. Learner language features have been examined by using keyword analysis by Jantunen (2011) who has been involved in collecting a corpus of learners' Finnish. Jantunen's keyword analysis is a good example of a corpus-driven study where the categories of keywords are defined in the process of examining the material. In addition, this study is an example of breaking the strict "rules" between corpus-based and corpus driven approaches: Jantunen uses annotated data for corpus-driven analysis.

It could be concluded that no matter what approach to corpora is used, corpus-based or corpus-driven, it is anyway impossible to keep intuition totally away from the study. For this reason, using annotation in a corpus-driven study could be justified.

References

Anthony, Laurence. 2005. AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. IWLeL 2004: An interactive workshop on language e-Learning: 7–13. https://dspace.wul.waseda.ac.jp/dspace/bitstream/2065/1390/1/01.pdf. Accessed 10 Aug 2014.

Anthony, Laurence. 2006. Developing a freeware, multiplatform corpus analysis toolkit for the technical writing classroom. *IEEE Transactions on Professional Communication* 49(3): 275–286.

Baker, Mona. 1993. Corpus linguistics and translation Studies. Implications and applications. In *Text and technology*, *honour of John Sinclair*, eds. Mona Baker, Gill Francis, and ElenaTognini-Bonelli, 233–250. Philadelphia and Amsterdam: John Benjamins.

Baker, Mona. 1995. Corpora in translation studies: An overview and some suggestions. *Target* 7(2): 223–243.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and translation. Studies in language engineering in honour of Juan C. Sager*, ed. Harold Somers, 175–187. Amsterdam: John Benjamins.

Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. *Target* 12(2): 241–266. doi:10.1075/target.12.2.

Baker, Paul. 2004. Querying keywords. Questions of difference, frequency, and sense in keyword analysis. *Journal of English Linguistics* 32(4):346–359.

Baker, Paul. 2006. *Using corpora in discourse analysis*. London and New York: Continuum.

Baker, Paul. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.

Becher, Viktor. 2010. Abandoning the notion of “translation-inherent” explicitation. Against a dogma of translation studies. *Across Languages and Cultures* 11(1): 1–28.

Berber Sardinha, Tony. 2004. *Linguistica de Corpus*. São Paulo: Manole.

Bernardini, Silvia, and Adriano Ferraresi. 2011. Practice, description and theory come together: Normalization or interference in Italian technical translation? *Meta* 56(2): 226–246.

Biber, Douglas, and Edward Finnegan, eds. 1994. *Sociolinguistic perspectives on register*. Oxford: Oxford University Press.

Culpeper, Jonathan. 2012. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeares’s Romeo and Juliet. In *Corpus linguistics. Volume IV. Methods and applications*, eds. Douglas Biber, and Randi Reppen, 81–109. London: SAGE.

Sutter, De, Patrick Goethals Gert, Torsten Leuschner, and Sonia Vandepitte. 2012. Towards methodologically more rigorous corpus-based translation studies. *Across Languages and Cultures* 13(2): 137–143.

Eckert, Penelope, and John R. Rickford. 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.

Firth, J. R. 1968 [1957]. A synopsis of linguistic theory, 1930–55. In *Selected papers of J.R. Firth 1952–59*, ed. Frank R. Palmer, 168–205. London: Longmans.

Firth, J.R. 1969 [1957]. The technique of semantics. In *Papers in linguistics 1934–1951,* 7–33. Reprinted. London: Oxford University Press.

Francis, Jill. 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In *Advances in Written Text Analysis*, ed. Malcolm Coulthard, 83–101. London: Routledge.

Gabrielatos, Costas. 2014. *Bibliography (in progress): Corpus approaches to discourse studies*. Latest update 15.09.2014. http://www.gabrielatos.com/CLDA-Biblio.htm. Accessed 15 Dec 2014.

Halliday, M.A.K. 1973. *Explorations in the functions of language*. London: Edward Arnold.

Halliday, M.A.K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M.A.K. 1994. *An introduction to functional grammar*, 2nd ed. London: Edward Arnold.

Halliday, M.A.K., and Ruqaiya Hasan. 1976. *Cohesion in english*. London: Longman.

Hoover, David L. 2003. Frequent collocations and authorial style. *Literary and Linquistic Computing* 18(3): 261–286.

House, Juliane. 2008. Beyond intervention: Universals in translation. trans-kom. *Zeitscrift für Translationswissenschaft und Fachkommunikation*. 1(1):6–19.

Jantunen, Jarmo H. 2011. Avainsana-analyysi annotoidun oppijakieliaineiston tutkimisessa: Alustavia havaintoja [Keyword analysis in examining an annotated learner language data: Preliminary findings]. In AFinLA-e. Soveltavan kielitieteen tutkimuksia, eds. Esa Lehtinen, Sirkku Aaltonen, Merja Koskela, Elina Nevasaari, and Mariann Skog-Södersved, n:o 3: 48–61.

Ji, Meng. 2010. General phraseological patterns in Yang’s translation. In: Phraseology in Corpus-Based Translation Studies. New Trends in Translation Studies. Vol. 1. Oxford et al.: Peter Lang.

Kemppanen, Hannu. 2004. Keywords and ideology in translated history texts: a corpus-based analysis. *Across Languages and Cultures* 5(1): 89–106.

Kemppanen, Hannu. 2008. Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi [Keywords and Ideology: A Corpus-based Analysis of Translated and Non-Translated History Texts]. University of Joensuu Publications in the Humanities 51. Joensuu: University of Joensuu.

Kemppanen, Hannu, and Jukka Mäkisalo. 2016. The foreign and the domestic in translations: combining reception and corpus analysis. In *Empirical translation studies. Interdisciplinary methodologies explored*, ed. Meng Ji. Sheffield and Bistol, 30–50: Equinox Publishing.

Kenny, Dorothy. 2001. Lexical Creativity in Translation. A Corpus-based Study. Manchester, UK and Northampton MA: St. Jerome Publishing.

Kuusinen, Martti, Veera Ollikainen, and Julia Syrjäläinen. 2007. Uusi venäläis-suomalainen suursanakirja. Kaksi osaa [New Russian-Finnish General Dictionary. In Two Parts]. Živoj Jazyk: Moskva.

Laviosa, Sara. 2002. Europe in the making in translational and-non translational English. In *CTIS occasional papers*, volume 2, ed. Keith Harvey, 85–95. Manchester: Centre of Translation and Intercultural Studies, UMIST.
Laviosa, Sara. 2003. Corpus-based translation studies: Where does it come from? Where is it going? A paper read at the conference Corpus-based Translation Studies Research and Applications Conference, 21–25 July 2003. Pretoria: University of South Africa.
Laviosa, Sara. 2011. Corpus-based translation studies: Where does it come from? Where is it going? In *Corpus-based translation studies. Research and applications*, eds. Alet Kruger, Kim Wallmach and Jeremy Munday, 13–32. London and New York: Continuum.
Maher, Ailish, Stephen Waller, and Mary Ellen Kerans. 2008. Acquiring or enhancing a translation specialism: the monolingual corpus-guided approach. *JosTrans, The Journal of Specialised Translation* 10.
Mahlberg, Michaela. 2005. *English general nouns: A corpus theoretical approach*. Amsterdam and Philadelphia: John Benjamins.
Mahlberg, Michaela. 2007a. Corpora and translation studies:Textual function of lexis in Bleak house and in a translation of the novel into German. In *La Trazuzione: Lo Stato dell'Arte. Translation: The State of Art*, eds. Vittoria Intonti, Graziella Todisco, and Maristella Gatto, 115–135. Ravenna: Longo.
Mahlberg, Michaela. 2007b. Corpus stylistics: Bridging the gap between linguistics and literary studies. In *Text, discourse and corpora*, eds. Michael Hoey, Michaela Mahlberg, Michael Stubbs, and Wolfgang Teubert, 219–246. London: Continuum.
Mahlberg, Michaela. 2009. Patterns in news stories: A corpus approach to teaching discourse analysis. In *Using corpora to learn about language and discourse*, ed. Linda Lombardo, 99–132. Berlin et al.: Peter Lang.
Mahlberg, Michaela. 2013. *Corpus stylistics and Dickens's fiction*. New York and London: Routledge.
Mauranen, Anna. 2000. Comparable corpus of history texts. A corpus compiled at the University of Joensuu within the research project Translated Finnish and Translation Universals. A Corpus-based Study (1997–2001).
McEnery, Tony, Richard Xiao, and Yukio Tono. 2006. *Corpus-Based Language Studies*. An Advanced Source Book. London and New York: Routledge.
McIntyre, Dan. 2007. Review of J. Sinclair (2004). Trust the text: Language, corpus and discourse. *International Journal of Corpus Linguistics*, 12(4): 563–575.
McNeill, Patrick. 1990. *Research methods*, 2nd ed. London: Routledge.
Mikhailov, Mikhail, and Miia Villikka. 2001. Is there such a thing as a translator's style? In Proceedings of Corpus Linguistics 2001, eds. Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, and Shereen Khoja, 378–384. Lanchaster: Lanchaster University. http://ucrel.lancs.ac.uk/publications/cl2003/CL2001%20conference/contents.htm. Accessed 15 December 2014.
Mulderrig, Jane. 2011a. Grammar of governance. *Critical Discourse Studies* 8(1): 45–68. doi:10.1080/17405904.2011.533570.
Mulderrig, Jane. 2011b. Manufacturing consent: A corpus-based critical discourse analysis of New Labour's educational governance. Educational Philosophy and Theory. *Special Issue: the Power in/of Language*, 43(6): 562–578.
Munday, Jeremy. 2008. Style and Ideology in Translation: Latin American Writing in English. London and New York.
Nevalainen, Sampo. 2004. Colloquialisms in translated text. Double illusion? *Across Languages and Cultures* 5(1): 67–88.
Newby, Howard. 1977. In the field: Reflections on the study of Sunfolk farm workers. In *Doing Sociological Research*, ed. Colin Bell, and Howard Newby, 108–129. London: Allen and Unwin.
Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London and New York: Routledge.

Ooi, Vincent B.Y. 2000. Asian or Western realities? Collocations in Singaporean-Malaysian English. In *Corpora Galore: Analyses and Techniques in Describing English*, ed. John M. Kirk, 73–89. Amsterdam: Rodopi.

Partington, Alan. 1998. *Patterns and meanings*. John Benjamins: Using Corpora in English Language Teaching. Amsterdam.

Pearson, Jennifer. 2000. Using specialized comparable corpora to evaluate student translations. In PALC': Practical Applications in Laguage Corpora- Papers from International Conference at the University of Łódź, 15–18 April 1999, eds. Barbara Lewandowska-Tomaszczyk, and Partick James Melia, 541–551. Łódź Studies in Language, vol. 1. Frankfurt am Main: Peter Lang.

Philips, Martin. 1989. Lexical Structure of Text. Discourse Analysis Monographs 12, Birmingham: University of Birmingham.

Rayson, Paul. 2012. From key words to key semantic domains. In Corpus Linguistics. Volume IV. Methods and Applications, eds. Douglas Biber and Randi Reppen, 173–202. Los Angeles et al.: SAGE.

Romagnuolo, Anna. 2009. Political discourse in translation: A Corpus-based perspective on Presidential Inaugurals. *Translation and Interpreting Studies Journal* 4(1): 1–30. doi:10.1075/tis.4.1.01rom.

Römer, Ute, and Stefanie Wulff. 2010. Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research* 2(2): 99–127.

Saldanha, Gabriela. 2011. Style of translation: The use of foreign words in translations by Margaret Jull Costa and Peter Bush. In *Corpus-based translation studies. Research and applications*, eds. Alet. Kruger, Kim Wallmach and Jeremy Munday, 237–258. London and New York: Continuum.

Scott, Mike. 2000. Focusing on the text and its key words. In *Rethinking language pedagogy from corpus perspective*, eds. Lou Burnard, and Tony McEnery, volume 2: 103–122. Frankfurt: Peter Lang.

Scott, Mike. 2009. In search of a bad reference corpus. In *What's in a Word-list? Investigating Word Frequency and Keyword Extraction*, ed. Dawn Archer, 79–92. Oxford: Ashgate.

Scott, Mike, and Christopher Tribble. 2006. *Textual Patterns*. John Benjamins: Key words and Corpus Analysis in Language Education. Amsterdam and Philadelphia.

Scott, Mike. 2014. WordSmith tools manual. Version 6.0. Oxford: Oxford University Press.

Semino, Elena, and Jonathan Culpeper. 2002. *Cognitive stylistics: Language and cognition in textual analysis*. London and Manchester: John Benjamins.

Stubbs, Michael. 1996. *Text and corpus analysis. Computer-assisted studies of language and culture*. Oxford: Blackwell.

Stubbs, Michael. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam and Philadelphia: John Benjamins.

Wierzbicka, Anna. 1997. *Understanding cultures through their key words*. Oxford: Oxford University Press.

Wilkinson, Michael. 2012. The best freeware corpus analysis program for translators? *Translation Journal*. 16(2).

Williams, Raymond. 1976. *Keywords: A vocabulary of culture and society*. London: Fontana/Croom Helm.

Winters, Marion. 2009. Modal particles explained: How modal particles creep into translations and reveal translators 'styles'. *Target* 21(1): 74–97.

Wynne, Martin. 2006. Stylistics: Corpus approaches. In Encyclopaedia of Linguistics, ed. Keith Brown, 223–226. 2nd ed. Oxford: Elsevier.

Xiao, Zhonghua, and Tony McEnery 2005. Two approaches to genre analysis. Three genres in modern American English. Journal of English Linguistics. Volume 33 (1): 62–82.

Zanettin, Federico. 2012. Translation-driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies. Manchester, UK and Kinderhook (NY), USA.

Zhang, Meifang, and Jeremy Munday. Forthcoming (eds.). Discourse Analysis and Translation. Special issue of Target. International Journal of Translation Studies. Amsterdam: John Benjamins.

Research materials

Ahmednabiev, M.K. 2007. Osobennosti formirovanija èkologičeskoj kul'tury studentov pedagogičeskih vuzov. Pedagogika 7, 61–64. http://www.portalus.ru/modules/shkola/rus_readme.php?subaction=showfull&id=1195049169&archive=1195938639&start_from=&ucat=&. Accessed 15 Aug 2014.

Bystrovskij, V.G. 2003. Predmet i principy semejnoj mediciny. Izdatel'skij dom. Russkij vrač. Stat'i iz žurnalov. http://www.rusvrach.ru/articles. Accessed 15 August 2014.

Mar'janenko, A.P. 2008. Iz čego sdelana korporacija "Rossija". NP "Prokonsul". Informacionnyj portal. http://proconsul-erg.ru/konsalt-org-busi-met-org.html. Accessed 15 August 2014.

Nezavisimaja gazeta. 1999. One year volume of the Russian newspaper Nezavisimaja gazeta. CD-rom. CDNG-98. Disk 1 (tekst). Moskva: Nezavisimaja gazeta.

Chapter 4
A Contextual Approach to Translation Equivalence

4.1 Introduction

In this chapter I shall be looking at fictional dialogue in a translation corpus as a case study in a contextual approach to translation equivalence informed by quantitative corpus evidence and sociocultural analysis.

The corpus is made up of novels written in English and their translations into Brazilian Portuguese, first carried out in the 1930–1950s and later retranslated in the 2000s. Its compilation was motivated by the fact that ample contextual information is available on the first translations, produced and published during the so-called Golden Age of Translation in Latin America, a historical period characterized by the emergence and growth of cultural industries propelled by a process of industrial growth and modernization that fostered the expansion of the educational system and a flourishing editorial market to cater for the needs of an emergent readership (Pagano 2001; Milton 2002). Studies focusing on the style of translations at that time have pointed to recurrent features that have been ascribed to strategies on the part of translators and editors to produce texts that could be more accessible and appealing to readers. Among those strategies, insertion of explanations, text abridgement and omission of foreign cultural references have been reported (Milton 2002). Another widespread strategy is connected to speech presentation in the novels whereby translators chose to render reporting verbs of the neutral type in the original texts (e.g., ‘say’, ‘ask’), by verbs that signal purpose, attitude, manner of speaking and discourse organization (Pagano 2002; Jesus and Pagano 2007). In contrast to the translations published in the 1930–1950 decades, retranslations in the 1990s and the 2000s have been found to make use of different strategies and to follow source text patterns more closely than first translations (Alves 2006).

Even though the reported research had a sound empirical basis and in some cases made use of corpus linguistic tools to query the texts, studies benefitted very little from computational resources nowadays available to us and which can greatly enhance our analysis. In this sense, methodological advances such as those offered

S. Laviosa et al., *Textual and Contextual Analysis in Empirical Translation Studies*, New Frontiers in Translation Studies,
DOI 10.1007/978-981-10-1969-2_4

by the software and computational environment R (R Core Team 2014) allow for revisiting claims made about the Golden Age of Translation and re-examining them within the framework of a comprehensive theory of language such as systemic-functional linguistics (Halliday and Matthiessen 2014).

With that aim in mind, a corpus-based study was designed in order to compare first translations and retranslations in terms of the equivalence relations established between source and target texts and gather insights into contextual and meta-contextual features of their production. The choice of the compiled corpus, additionally, made it suitable to investigate the so-called retranslation hypothesis (Berman 1990), which posits stages in the cultural accommodation of a literary work of art in a target culture. First translations are posited as a first step towards introducing a foreign work with the implication that they need to follow target language and culture patterns more closely than subsequent translations. Thus, for the present study computational statistics were selected to measure the degree of proximity between originals, first translations and retranslations in an attempt to probe the retranslation hypothesis through an empirical approach.

Drawing on systemic functional theory and accounts of translation (Catford 1965; Halliday [1960] 2005a, 1992a; Matthiessen 2001) and incorporating insights from narratology (Toolan 2001, 2009) and literary stylistics (Leech and Short 1981; Semino and Short 2004), a study of verbs serving in verbal clauses used to quote and report in fictional dialogue was carried out based on data from a bilingual parallel corpus of novels originally written in English and their translations and retranslations into Brazilian Portuguese. The aim was to examine textual equivalents found in the corpus in terms of formal correspondence and shifts in order to identify patterns in the use of verbs and interpret them in the light of the context and the meta-context of the translation task.

Clause-aligned samples of dialogues retrieved from the novels were manually annotated for grammatical functions, lexis, narrative level and type of equivalence holding between source and target text clauses. A database was generated in R (R Core Team 2014), where frequency counts for grammatical functions and equivalence types were obtained and hierarchical cluster analysis was performed in order to measure proximity of patterns between the originals, first translations and retranslations. Frequency counts of lexical verbs were also obtained and examined in conjunction with the cluster analysis results.

This chapter will report the main findings from the study regarding patterns identified and what they reveal about the context and meta-context within which first translations and retranslations were produced. The findings are also discussed in the light of the retranslation hypothesis and the claims made regarding translation in the 1930–1950s as contrasted with retranslations in the 2000s.

The following sections after this introduction will provide a brief overview of the theories and concepts that were drawn upon for the design of the present study as well as previous findings in the literature that will frame the discussion of the results herein reported. Where suitable, examples retrieved from the corpus analysed in this study will be used for the purposes of illustration of theoretical concepts. Throughout the chapter examples retrieved from the Brazilian Portuguese texts will be accompanied by an English gloss underneath.

4.2 Contextualizing Translation

A pending issue in translation studies is how to operationalize a theory-informed notion of context that allows for integrating text and sociocultural analysis. Among the most fruitful attempts are studies drawing on systemic functional theory (SFT) which, as a comprehensive theory of language, has offered a model in order to contextualize translation from the most global environment within which it can be mapped (context of culture) to the most local one (grammar), postulating a meta-context for the translation task itself (Matthiessen 2001). A tenet in SFT is the reliance on corpus for exploring language in quantitative terms, thus providing an empirical grounding for whatever claims that may be made in any of the environments focused upon in our analysis.

Sharing the assumption that the corpus is an essential piece in any attempt at describing and explaining translational phenomena, corpus-based translation studies (CTS) has drawn on SFT as a theoretical framework to interpret results obtained from the corpus by reference to the multiple systems at play for meaning making in the language system (Munday 2002). Most significantly, SFT offers CTS a framework to formulate potentially insightful questions to query the corpus.

SFT conceives of language as a stratified system (Halliday and Matthiessen 2014). Content and expression are the two main strata, the former subdivided into semantics and lexicogrammar and the latter into phonology and phonetics. Each stratum is realized by the stratum lower in hierarchy. Thus, semantics is realized by lexicogrammar, in turn, realized by phonology and ultimately phonetics.

For SFT, language is a potential that is instantiated in text. When analyzing text we are dealing with particular resources of the system that have been instantiated in a particular specimen. Every text clusters with similar texts, which make a text type. We identify text types because they tend to share similar characteristics and we produce texts of a particular type by means of the resources of the language system that we know are critical for that particular type. These resources are studied in SFT as registers or subspecifications of the language system.

The language system is embedded in context, which can be specified for the different points along the cline of instantiation: system—register/text type—text. The "contextual potential of a community" (Halliday and Matthiessen 2014, 32) or context of culture comprises its potential to make meanings through language and other semiotic systems. The cultural potential is made up by different domains or institutions, themselves made up by particular situations which can be clustered into situation types. The language used to make meanings in the different situation types can be studied by making abstractions from the texts therein produced.

A situation involves some socio-semiotic activity in which language takes a part. One such activity is, for instance, recreating events of the human world or a fantasy world through narratives and dramatizations. This is done through texts such as narratives, tales, novels and theatre plays. Among the language resources recruited by recreating situations are, for instance, figures of speech in the semantics stratum, typically realized in the lexicogrammatical stratum by verbal clauses, where lexical verbs of saying are typically used. Thus, we can study a particular lexicogrammatical selection by reference to the text type where it typically occurs and the selections that are made in the language system to generate that particular text type.

This movement of successively looking at different points in the cline of context where the language used to produce a meaning can be sketched is a powerful principle in SFT-based analysis and one that is enlightening for an attempt at contextualizing translation.

In that sense, Matthiessen (2001) proposes an account of translation through the location of the different contextual environments where we can examine texts instantiating meanings in different language systems. This can be done by working with particular texts, bearing in mind that they may be approached as specimens of text-types that can reveal prototypical choices made in the language system. These environments can be mapped by different coordinates in instantiation and stratification, that is, by looking at the different points along the cline from text to system and the realizations of semantics by the lexicogrammar.

Matthiessen proposes modeling the context of the translation task as a meta-context, being "more abstract than what is being translated" (111) and defined by variables such as the drafting and editing of the translated text, the relationship between the author of the original text, the translator, the editor and the reader, the guidelines in the commissioning of the translation, and the role envisaged for the translated text in the target culture, including "the context in which the translation is read or listened to" (111). Contextualizing translation thus involves examining both the context of source and target texts and the meta-context of translation.

Halliday (1992b, 15) defines translation as "*guided* creation of meaning" and emphasizes the conditioning role that contextual parameters play in that particular process of meaning making: the translator needs to abstract a context out of the source text and then use it as a guide to create a target text, guided, in turn, by a target text context, which the translator constructs bearing in mind prospective readers. The "ideal" translation would then be "that in which the form in Language 2 operates with identical effect in the identical context of situation as the form in Language 1" (Halliday 2005b, 9). Equivalence, as we saw, is equivalence of context and it is the task of the linguist, as Halliday explains, to model that process trying to locate the environment within which equivalence can be most adequately operate.

There are several implications to this. One is that we can abstract relevant features of the context at which equivalence operates and learn about the translator's construction of such context, which informs us about how the translator abstracted contextual parameters from the source text and posited parameters for the target text. Another one, and a particular insightful for the purposes of the present study, is that actual and potential choices are equally relevant, that is, what the translator chose and what he or she could have chosen instead. In the case of multiple translations of the same source text, as is the case of the study herein reported, the different target texts represent for the analyst an actualization of potential choices and are in that sense highly revealing about translation contexts.

Since the study of translation, as Matthiessen (2001) remarks, proceeds from instance to system in the cline of instantiation, the starting point for a SFT informed analysis of a translational corpus are text equivalents, as defined in the following section.

4.3 Equivalence and Shifts in Context

One of the most revisited models of translation equivalence in the discipline of translation studies is Catford's (1965), based on an early version of Halliday's systemic-functional grammar and later expanded by Matthiessen (2001) to incorporate further developments in the theory from the 1960s to the present.

From a systemic-functional perspective, as Halliday himself puts it (1992a, 16), translation has to do with "equivalence of function in context", which has implications both for the task of the translator who produces a translation and the task of the analyst of a translated text. The translator makes decisions about the best environment to achieve functional equivalence and the analyst examines translation equivalents in order to verify which contextual parameters can best account for the decisions made by the translator.

Catford posited translation equivalence as an empirical phenomenon and suggested it could be studied by identifying textual equivalents in a corpus and analyzing how they stand in relation to one another in terms of formal correspondence, when they involve similar choices made in analogous grammatical systems, or shift, when different choices were made in different grammatical systems in the source and target languages. Assuming semantic equivalence, Catford focused on the stratum of lexicogrammar and the rank scale in particular to define correspondence or shift. Table 4.1 illustrates textual equivalents standing in a relationship of formal correspondence:

Table 4.1 Instance of text equivalence: formal correspondence

Source text	Your wife's lovely
Target text	Sua esposa é encantadora
gloss	*Your wife is charming*

As Table 4.2 below shows, correspondence means that similar choices are made in each language system starting at the clause level and moving down the rank scale to the group and word levels, the analysis potentially reaching the morpheme level, if so wished. In our example, we have a single clause (a finite clause) made up of three groups (two noun groups and a verbal group), two of them being one-word groups and one a two-word group.

Structurally, the clauses in Table 4.2 are made up by groups following an analogous sequence and the groups are also structured in the same sequence of components.

Table 4.2 Rank scale analysis of translation equivalents in relation of formal correspondence

Source text	Clause	Your wife is lovely			
	Group	Your wife		is	Lovely
	Word	Your	Wife	is	Lovely
Target text	*gloss*	*Your*	*Wife*	*is*	*Charming*
	Word	Sua	esposa	é	encantadora
	Group	Sua esposa		é	encantadora
	Clause	Sua esposa é encantadora			

When formal correspondence cannot be assigned to text equivalents, because they do not share similar choices in the grammatical systems chosen for comparison, a shift is said to have taken place. Table 4.3 illustrates this:

Table 4.3 Instance of text equivalence: shift

Source text	Clause	A marvellous stillness pervaded the world					
	Group	A marvellous stillness		pervaded	the world		
	Word	A	marvellous	stillness	pervaded	the	world
Target text	*gloss*	*Invaded*	*the*	*world*	*a*	*peace*	*marvellous*
	word	Invadia	o	mundo	uma	paz	maravilhosa
	group	Invadia		o mundo	uma paz maravilhosa		
	clause	Invadia o mundo uma paz maravilhosa					

There are two major shifts in the example displayed in Table 4.3. The clause structure in English is made up of three functions, realized by a noun group, a verbal group and a noun group in that sequence. In Portuguese, the clause structure follows a different sequence, the first function realized by a verbal group, the second by a noun group and the third one by a noun group. The sequence of components in the English noun group "a marvellous stillness" is also different from the equivalent noun group in Portuguese "uma paz maravilhosa" ("a peace marvellous"). The difference in the sequence of functions between the English and the Portuguese clause is a structure shift in Catford's terms.

Matthiessen (2001) broadens the scope in the mapping of equivalence originally envisaged by Catford, to contemplate the different strata in the language system (lexicogrammar, semantics and context) and other dimensions in the organization of language as developed by Halliday from the 1960s onward. Matthiessen points to two dimensions in particular as being the most implicated in translation shifts. These are structure, already dealt with by Catford, and system. By structure, as we saw in the examples above, we refer to the organization of language in layers ordered by the principle of rank, whereby there is hierarchical relation running from the clause to the morpheme with the group and the word as intermediary levels. There is a configuration of functions in each unit at each rank. By system, we refer to the organization of language in paradigms of available options which can be further and further specified along a network. This principle of progressive sub-specification is called "delicacy".

Shifts in systems can involve both grammar and lexis. In SFL, lexis is actually grammar seen from a more delicate perspective. The example in Table 4.4 illustrates a shift within the grammatical system of mood. Both the source text and target text clauses share the same mood type: indicative and the same indicative type: interrogative. However, when it comes to the type of interrogative, the source clause is a Yes/No question, whereas the Target text clause is a Wh-question.

Table 4.4 Instance of text equivalence: shift in grammar

Source text	"Do you suppose he killed her?" Nora asked
Target text	Quem imagina que possa ter matado essa moça?—perguntou Nora
gloss	*Who do you think might have killed that girl? asked Nora*

Shifts at more delicate grammar are also very frequent and may involve choices in different taxonomies of lexical words. The example in Table 4.5 shows one such shift.

Table 4.5 Instance of text equivalence: shift in lexis

Source text	Dorothy said: "Damn it."
Target text	- Droga!—exclamou Dorothy
gloss	- *Damn it!—exclaimed Dorothy*

In the source text, "said" is a neutral, general verb of saying used for quoting in fictional dialogues. In the target text, this is rendered by a verb specifying manner of saying: "exclaim" ("exclamar").

As Matthiessen remarks, shifts in grammatical systems tend to have a greater impact on meaning construal than those at more delicate grammar. The examples in Table 4.4 illustrate this. The source text Yes/No question seeks for confirmation about a particular piece of known information regarding the perpetrator of a murder ("he"); the target text Wh-question seeks for a completely unknown, missing piece of information ("Who") regarding the murderer.

The criteria by which we ascribe correspondence or shift to text equivalents is largely dependent upon the particular level of granularity we opt for in our analysis, that is, how far in delicacy we want to advance when examining grammatical systems and how low in the rank scale we want to move in structure. Shifts at one level may be part of formal correspondence at a higher one. An explanation for this is provided by Matthiessen (2001), who works out the mechanism whereby shifts operate within conditions of translational equivalence. According to the author, a "higher level constant" (74) or relationship of equivalence must always obtain at a higher point in the language system, so that shifts may take place at a lower level in the strata or rank scale or further ahead in a system network. From a dynamic perspective of the textual process, Halliday (1992b, 22) argues that every single choice a translator makes at a local level or focal point in a translation task needs to be in consonance with the translator's overall view of the task as a whole. Based on this assumption, the analyst focuses on a particular item in the text and starts by looking at the immediate grammatical environment to interpret the concept of equivalence. When the local context does not provide a basis for interpreting a shift, the analyst needs to go beyond the stratum of lexicogrammar and semantics and search for a likely explanation in the context of culture and the meta-context of translation.

In analyzing a case such as the example in Table 4.6, going beyond the immediate grammatical environment is necessary to see how each clause operates in each text.

Table 4.6 Instance of text equivalence with shift in lexis due to cohesive co-reference

Source text	"What are you going to do with it?" Nora asked
Target text	– Que vai fazer com esse revolver, Nick?
gloss	*What are you going to do with that revolver, Nick?*

The English clause makes use of a pronoun ("it"), whose referent needs to be retrieved anaphorically in the clauses preceding that dialogic exchange to realize it is a weapon that the narrator, Nick, is examining in the presence of his wife, Nora.

The Portuguese clause makes that meaning explicit through the lexical noun "revolver", a synonym for the word "arma" ("weapon") appearing in the preceding clauses.

This example shows how the system of cohesion operates differently in the two language systems regarding the system of personal reference. In English, the system is tripartite (feminine, masculine and neuter) and the pronoun "it" can only refer to the weapon mentioned in the preceding discourse. Thus it is easily retrievable by the reader. In Portuguese, the system of personal reference is bipartite (feminine and masculine). A pronoun to refer back to the noun "arma" ("weapon") would be a feminine pronoun ["ela" ("her")], which could bring about ambiguity as to whether Nora is referring to the weapon or to the girl who brought the weapon and is asleep in their apartment. In this sense, the choice of a lexical noun avoids ambiguity. Additionally, the choice of the noun "revolver", masculine in the system of personal reference, makes it clearer that the following clauses will refer back to it through the pronoun "ele" ("he") and thus avoid any ambiguity with the girl.

Another shift we can see in this example is the non-realization of the clause "Nora asked" in the Portuguese rendition. This has to do with narrative progression and contextual features. The dialogue where this exchange takes place has two main interlocutors—Nick, narrator and character, and his wife Nora. The preceding discourse allows the reader to infer that the question about the weapon is posed by Nora to the narrator. Even so, the English text makes use of a clause ("Nora asked"), most probably because the question is the first turn in an exchange after a narrative passage in the novel. In Portuguese, there is no introductory clause explicitating who posed the question, a not-at-all problematic choice, since like in English, the identity of the speaker can be easily inferred by the reader from the immediate co-text. Such a shift in the Portuguese clause may be ascribed to preferences in the context of situation and culture for this kind of text in each culture—this is a dialogic exchange in a fictional narrative, a kind of text related to the socio-semiotic activity of recreating an imaginary world. It may be ascribed, too, to meta-contextual parameters of the translation, in the sense that it may be connected to translator's decisions regarding how the source text is going to be presented in the target culture and which role is envisaged for the prospective target reader of the text, including perhaps features related to the commissioning of the translation by a literary agent, such as publishers.

Equivalence, as Matthiessen notes, generally operates at the stratum of semantics and at the level of the clause in the rank scale. Nevertheless, there might be shifts like in Tables 4.4 and 4.6, which can only be accounted for in terms of contextual features of the text and the task of translation. In such cases, granted that equivalence obtains at the level of context, the shifts may be located, as we saw, at the stratum of semantics with repercussions on lexicogrammar.

The identification of shifts in a translational corpus can be a very time consuming task, little assisted by the computer, and clearly requiring close human analysis and manual annotation of lexicogrammatical functions. This hard enterprise is worth pursuing if shifts identified locally can be related to a global pattern in text to be accounted for by resorting to contextual analysis. Halliday (2005b) proposes a compromise between human and computer labor by concentrating our manual annotation on those grammatical systems that are critical to theorize a phenomenon under study. This solves in part the problem, but still the extent manual annotation can reach will never compete with the amount of data a computer can process automatically. In this sense, working on small and representative samples of a corpus offers a viable way of enhancing empirical studies at a high level of granularity and with the support of quantitative data.

Counting and classifying shifts in manually annotated representative samples of a corpus is the proposal here advanced with a view to identifying patterns and interpreting these in the light of results obtained in sociocultural analysis of a translational corpus. This is an emergent approach in the literature on empirical translation studies together with other statistical methods which have brought a quantitative dimension to corpus investigation allied with contextual analysis (Oakes and Ji 2012; Ji 2014). The next sections will present the object of study in our research—verbal clauses used in one of the modes of speech presentation in novels—and the grammatical and narratological categories annotated in our study. Reported research in the literature will be briefly presented as well as the sociocultural background against which the results of the study will be interpreted.

4.4 Speech Presentation

Among the objects of inquiry of stylistic studies of the translated text is discourse presentation in fictional narrative and the challenges it poses to the translator. Discourse presentation refers to the modes in which the speech, thought and writing of narrators and characters are presented. These modes are usually located along scales as proposed by Leech and Short (1981) and Semino and Short (2004).

From the perspective of systemic-functional theory and regarding speech presentation in particular, which is the object of inquiry in the study herein reported, meanings related to the domain of experience of saying are construed through verbal clauses. These have a grammatical configuration of functions labelled Participants, Processes and Circumstances. Participants include the producer of a verbal sign or saying (Participant Sayer); that which is said by the Sayer or Verbiage; and, if the Sayer addresses another participant, the addressee or Receiver. When there is a Participant affected by the process of saying, that is, a Participant impacted by means of a verbal Process, this is called Target. Participants are grammatically realized by nominal groups. Processes in a verbal clause are realized by verbal groups having verbs of saying, also known in the literature as reporting verbs or *verba dicendi*. Tables 4.7, 4.8 and 4.9 show examples illustrating different configurations of verbal clauses with the main grammatical functions labelled.

Table 4.7 Verbal clause with Sayer and Verbiage

They	told	their story
Participant Sayer	Verbal process	Participant Verbiage

Table 4.8 Verbal clause with Sayer, Receiver and Verbiage

Wynant	had said	to me	something about women and dogs
Participant Sayer	Verbal process	Participant Receiver	Participant Verbiage

Table 4.9 Verbal clause with Sayer and Target

Rocky	would announce	them
Participant Sayer	Verbal process	Participant Target

Verbal clauses may involve an act of saying as an activity of talking or as semiosis, i.e., a record of a speech event. In the latter, what is recorded may be the content of what is said (the Verbiage) or the projection of what someone said either by quoting or reporting their words. To quote or to report other people's words, a clause complex is usually used, in which a quoting/reporting verbal clause projects a quoted/reported clause. These two possibilities are termed direct speech in the case of quoting and indirect speech in the case of reporting. Examples in Tables 4.10, 4.11, 4.12 and 4.13 illustrate different acts of saying.

Table 4.10 Saying as activity of talking

I	didn't speak
Participant Sayer	Verbal process

Table 4.11 Saying as semiosis with Verbiage

You	said	nothing
Participant Sayer	Verbal process	Participant Verbiage

Table 4.12 Saying as semiosis projecting quoting

Mimi	said	'She's lying, I bet.'
Participant Sayer	Verbal process	Quote

Table 4.13 Saying as semiosis projecting reporting

Mimi	said	Asta was a lovely dog
Participant Sayer	Verbal process	Report

As already mentioned, the act of saying may involve reception, when there is a Receiver, or no reception, illustrated by the examples in Tables 4.14 and 4.15.

Table 4.14 Saying as semiosis projecting reporting—reception

He	asked	me	how the world was treating me
Participant Sayer	Verbal process	Participant Receiver	Report

Table 4.15 Saying as semiosis projecting quoting—non-reception

She	asked:	'Don't you?'
Participant Sayer	Verbal process	Quote

From a grammatical perspective, quoting and reporting involve a configuration of clause complexes made up by a quoting or reporting clause and a quoted and reported one. The former is usually referred to as "reporting clause" both for quoting and reporting, while the second clause is termed a "quote" or a "report". Tables 4.16 and 4.17 illustrate this with examples.

Table 4.16 Quoting

"That's my business,"	he said
Quoted clause	Quoting clause

Table 4.17 Reporting

I told Gloria	what James had said
Reporting clause	Reported clause

What is quoted or reported can be a proposition in an information exchange, i.e., a statement or a question, or a proposal in a goods-and-services exchange, i.e., an offer or a command. Tables 4.18, 4.19, 4.20 and 4.21 illustrate each configuration.

Table 4.18 Quoting proposition

"I didn't think,"	I said
Quoted clause	Quoting clause

Table 4.19 Reporting proposition

Dorothy sulkily repeated	that she didn't see why she had to waste an afternoon at Aunt Alice's
Reporting clause	Reported clause

Table 4.20 Quoting proposal

"Let him wait,"	he said
Quoted clause	Quoting clause

Table 4.21 Reporting proposal

I told him	to call his office
Reporting clause	Reported clause

Halliday and Matthiessen classify lexical verbs typically used in verbal clauses according to whether (i) they are general, unmarked choices, such as "say"; (ii) they specify a speech function (giving or demanding information or goods-and-services); or (iii) they elaborate on a speech function, adding a circumstantial feature to the process of saying or a specific manner or connotation. Lexical verbs typically used for each type are illustrated in Table 4.22.

Table 4.22 Examples of lexical verbs for each type of verb of saying

Example	Type of Verb of Saying
'A public wedding isn't so bad,' Ruby said	General member of class of verbs of saying ("say")
"Try his lawyer," I suggested	Verb specific to speech function (giving)
'How do you feel?' I asked	Verb specific to speech function (demanding)
"It's not that," I explained	Verb with additional circumstantial feature
"You said nothing," I whispered'	Verb specifying manner

The variety and subspecification of verbs of saying evidence the diversity of the grammatical resources available in the language system to represent our experience of speech events, highlighting: what an act of saying is used for ("ask", "suggest", "offer", demand"); what turn it is being taken ("add", "interrupt', "conclude"); the manner in which something is said ("insist", "whisper", "shout", "thunder"); the means of communication ("announce"), among several other features. The neutral verb "say" can also function to specify an act of saying when accompanied, for instance, by a circumstance of manner ("say sarcastically").

According to Halliday and Matthiessen, verbal clauses are typically used in narratives to construct dialogues which recreate interactions between narrator and characters or characters among each other. As we saw above, the quoting/reporting clause signals who is speaking while the quote or report is the content of what is said. Direct speech, i.e., quoting verbatim someone's words, is the norm for speech presentation in fictional narratives, probably because, Halliday and Matthiessen (2014, 528) argue, it adds a "more immediate and lifelike" effect to the dialogue.

The so-called reporting clause may be present or not. In the following dialogue retrieved from our corpus, each quote is accompanied by a reporting clause, explicitly signalling who says what. The reporting clauses are underlined for ease of visualization.

> 'What's the matter with Gloria?' <u>James asked me one day</u> as we came back to the floor from the sleeping quarters.
>
> 'Nothing. What do you mean?' <u>I asked</u>. But I knew what he meant. Gloria had been singing the blues again.
>
> 'She keeps telling Ruby what a chump she would be to have the baby,' <u>he said</u>. 'Gloria wants her to have an abortion.'
>
> 'I can't understand Gloria talking like that,' <u>I said</u>, trying to smooth things over.
>
> 'You tell her to lay off Ruby,' <u>he said</u>.
>
> When the whistle started us off on the 216th hour I told Gloria what James had said.
>
> 'Nuts to him,' <u>she said</u>. 'What does he know about it?'McCoy ([1935] 1995, 33)

Unlike the dialogue above, the following exchange has an initial reporting clause ("Nora asked") and then proceeds with a to-and-fro exchange between the narrator (a first person narrator) and the character (Nora) with no explicit indication of speaking voices till halfway through the conversation, when the narrator introduces his own speech with a reporting clause ("I said").

> "Do you suppose he killed her?" Nora asked when I put the paper down again.
>
> "Wynant? I wouldn't be surprised. He's batty as hell."
>
> "Did you know her?"
>
> "Yes. How about a drop of something to cut the phlegm?"
>
> "What was she like?"
>
> "Not bad," I said – "She wasn't bad-looking and she had a lot of sense and a lot of nerve – and it took both to live with that guy."
>
> "She lived with him?"
>
> "Yes. I want a drink, please. That is, it was like that when I knew them."
>
> "Why don't you have some breakfast first? Was she in love with him or was it just business?"
>
> "I don't know. It's too early for breakfast."
>
> (Hammett [1933] 1992a, 16)

The need for a reporting clause to be present in a dialogic exchange is a matter of controversy. On the one hand, Halliday and Matthiessen (2014) argue that speech introducing clauses are not necessary for narrative dialogue to take place, their occurrence being absolutely necessary only when their absence would prevent a reader from following the discourse flow. On the other hand, the amount and location of reporting clauses in a dialogic exchange are, according to the authors, a relevant variable in discourse production, which merits attention by text analysts.

Another feature of dramatic dialogue in fiction pointed out by Halliday and Matthiessen (2014) are quotes introduced by verbs which are not strictly verbs of saying and refer to a particular behaviour accompanying an act of speaking, sometimes evoking sounds typically associated to animals, as the examples in Table 4.23 illustrate.

Table 4.23 Examples of lexical verbs signaling behaviour

'What's the big idea?' Socks roared
'I wasn't expecting you, Lieutenant,' he whined

Stylistic studies of speech presentation in fiction have pointed out features of the use of direct speech in dialogue construction in novels and short stories. Building on systemic functional theory, Leech and Short (1981) argue that direct speech is the norm for speech presentation in fiction, a claim confirmed by Semino and Short's (2004) corpus analysis. Semino and Short also confirm Halliday and Matthiessen's remark as to "say" being the most frequently used verb for all modes of speech presentation and the significant number of occurrences of reporting verbs signalling behavior or manner of speaking, with a higher variety of reporting verbs in fiction when compared to non-fiction (96). Fiction is confirmed by the authors as having a higher amount of direct speech than other text types (97), direct speech being the most frequent mode of speech presentation (177). Semino and Short stress the "vividness, immediacy and involvement" of direct speech in narratives and interpret shifts from direct to indirect speech as a means of indicating a narrator's or a character's likely reluctance to further engage in conversation (80) or producing a distancing effect due to the contrast with the dramatic immediacy of direct speech (85). Toolan (2009) also points to the high frequency of direct speech in fictional narrative and stresses its role as an important resource to signal that the narrator is not responsible for the reported utterance (75). In effect, as Leech and Short's model posits, direct speech is located within a cline ranging from the loosest to the tightest control purportedly exercised by the narrator over the report of facts. Free Direct Speech, i.e., quotes with no reporting clause, is located closer at the extreme position of the cline, signalling that the narrator has no control over the report of facts; Direct Speech, i.e., quotes with accompanying reporting clauses, is located right next to it on the cline, with the implication that there is partial control by the narrator. Hence, the use of reporting clauses is not only connected to the need for identifying who says what, but also with the degree of narrator's interference or control. This explains why reporting clauses, particularly those with neutral verbs of saying, tend to recur in a dialogue, even if there is no need for that, since the interlocutors are clearly identified. The presence of a reporting clause signals mediation and distance, whereas its absence is associated with a faster and more spontaneous exchange.

Toolan explores direct speech in narrative progression by looking at cohesive chains and their role in advancing the story and generating expectations in the reader. According to the author, dialogues in narrative may be 'completive sequels', i.e. a sequence of speech moves which are initiated and concluded one by one or "prospection triggers', when a speech move is initiated but left open or unresolved. The latter are critical points in a narrative and demand skills from readers so that they may build a coherent narrative.

As regards the omission of reporting clauses, which in literary stylistics receives the name of free direct speech, Toolan (2009) accounts for this in terms of their not being particularly relevant to narrative progression, especially when the reporting verb is the general verb "say". The choice of reporting verbs other than "say" becomes relevant precisely due to their contrast with the more neutral ones (88).

Semino and Short characterize reporting clause omission as having the effect of speeding up the flow of a dialogic exchange (90). However, the authors also note, reporting clause omission involves a higher level of difficulty for readers to keep track of who says what in a dialogue (92). In such cases, readers need to rely on contextual clues to assign each utterance in an exchange to a particular character. This is confirmed by empirical studies in reading and cognitive stylistics (Emmot 1994; Emmot et al. 2007) that examine how readers need to build contextual information at hand when reading a narrative in order to interpret/assign identities to referents in order to track participants. These studies investigate reader response to a narrative as it unfolds with a view to identifying clues in the text that are assumed to influence how readers interpret the narrative. One such clue is speech and thought presentation and the presence of reporting clauses clearly stating who said what.

Reading studies (Elson 2012) have shown that competent readers manage to deal with the attribution of quotes to different characters in a dialogic sequence even when only the first quote is explicitly ascribed to an interlocutor. As they read, readers seem to build a discourse model encompassing the speakers currently interacting and assign quotes to alternating speakers in a sequence. Less competent readers and young children are reported to find this task more difficult (Weber 2008; Tompkins et al. 2013). This has also led to positing a model of reading and writing competence which attempts to measure a writer's "sense of audience" (Gregg et al. 1996), that is, how much and in what ways a writer produces a text tailored to an envisaged reader.

Text and discourse analysts have also pointed relevant features of fictional dialogue. Gregory (1967) draws a comprehensive classification of language varieties within the spoken-written cline and characterizes the language of fictional dialogues as "written to be read as if heard" (193). This accounts for the fact that authors sometimes include phonetic and phonological features of the characters' speech and make use of reporting verbs with added connotations, in what Gregory terms "invitations to an auditory experience" (193). Dialogues apart, the default mode for narrative is "written to be read 'in the sense of not read aloud'".

Drawing on a comparison of natural, spontaneous conversation and fictional dialogue, Tannen (1986) stresses the latter as a constructed artifice to promote reader's involvement in a narrative. She accounts for the presence of reporting clauses or their omission ("unintroduced dialogue") in terms of the writer's likely intention and the prospective reader's response. Thus omission tends to occur when reporting clauses have neutral verbs like "say" and the identities of the speakers can be easily tracked along the dialogue. An interesting feature pointed out by Tannen is that the use of reporting clauses with "say" may be sometimes restricted to one of the characters taking part in the interaction, the writer using reporting clauses in order to "keep the character in focus as the protagonist" (320). Reporting clauses with verbs other than "say" and neutral ones are posited as an evaluative device in the narrative (322), as a means of enhancing characteristics of the characters.

The particular nature of dialogue in fictional narrative, especially in novels and short stories, is ascribed by Halliday and Matthiessen (2014) to the fact that fictional narrative is a second-order socio-semiotic process, through which first-order experience, as is the case of natural conversation, is recreated for specific purposes. Their analysis requires that several contextual levels be taken into account. Thus taking for granted the supposed interaction between an implied author and an implied reader of a fictional narrative, there is a first level in which a narrator or a character in the narrative interacts with other characters, and there may still be a second or third level in which characters quote or report conversations of other characters. Narratology studies have extensively explored narrative level in connection with the narrator's participation in the story and the reliability of his/her narration (Rimmon-Kennan 2005).

Multiple narrative levels are illustrated in the following example in Table 4.24.

Table 4.24 Narrative levels

'Mr Donald told us his lawyer said the city couldn't do anything,' I said	Level 1	Narrator ("I") quotes a proposition
'Mr Donald told us his lawyer said the city couldn't do anything,'	Level 2	Character ("Mr Donald") reports a proposition to narrator + characters ("us")
His lawyer said the city couldn't do anything	Level 3	Character ("lawyer") reports a proposition

With regard to the relevance of fictional dialogue to literary genres in particular, Semino and Short (2004) report a higher number of speech presentation instances in their corpus for texts classified as popular fiction when contrasted with the so called classic novels in English. Crime and detective novels have been pointed out to stand out due to the frequency with which dialogue occurs and the role it plays in narrative development (De Haan 1996; Axelsson 2011). Besides the suggested uses for dramatic dialogue in scriptwriting textbooks (Field 1994), such as story advancement, character construction, scene to scene connection and verisimilitude, crime fiction seems to resort to dialogues as a mode through which the reader is invited to be part of a case or mystery investigation and resolution. Bönnemark (1997) (as reported in Axelsson 2011, 13) points to the role of dialogue as moments in which summaries and evaluation of cases are carried out, the final resolution frequently being arrived at as a result of a dialogic interaction. The significance of dialogues for texts functioning in our culture for recreating a socio-semiotic activity thus accounts for their relevant place in stylistic studies in general as well as in translation studies, as the following section will show.

4.5 Speech Presentation in Translation

The topic of speech presentation is a much researched topic in translation studies, particularly in corpus based studies of both bilingual and monolingual corpora. Shifts in modes of speech presentation and choice of reporting verbs are two of the issues frequently discussed by translation studies scholars.

Baker (2000) analyses a monolingual corpus of fiction translated into English by two translators and finds different patterns in their use of speech presentation modes and choice of reporting verbs. Baker focuses exclusively on translated text and does not look at the originals; however, her discussion ends with some remarks on the comparison between source and target texts, which evidences shifts in the use of reporting verbs by one of the translators. These are hypothesized to be due to decisions (whether conscious or not) prompted by the translator's view of the relationships between source and target culture and the prospective reader of the target texts.

Puurtinen (2003) analyses a corpus of translated and non-translated children's books in Finnish and finds that reporting verbs are more varied in translations than in non-translations. Pressed by the need to render the repeated use of "say" and other common verbs in English, Finnish translators are found to resort to a wider variety of reporting verbs than those found in Finnish non-translated texts. Like Baker (2000), Puurtinen does not compare source and target texts but points to the need for this comparison to be carried out so as to complement findings from a monolingual comparable corpus.

Working from a bilingual parallel corpus perspective, Bourne (2002) reports on an analysis of the Spanish translation of an English novel and finds that a wider variety of verbs are used to render “say”, which has an impact on the meanings construed in the target text. Likewise, Winters (2007) compares reporting verbs in two translations of Scott Fitzgerald simultaneously published in Germany and finds that one of the translators tends to use more repetition of lexical verbs, a pattern that is closer to the one in the source text in English, while the other uses a wider variety of verbs in German in order to avoid repetition, with the concomitant effect of adding connotations to the speech act that are not present in the source text. Winters mentions a personal communication in which the translator who uses a wider variety of verbs accounts for her choice in terms of her assumption that repetition is bad style in German.

Similar patterns of reporting verbs in translated text are found by Cruz (2003), who compared an English novel translated into Brazilian Portuguese. Her results attest to the tendency for translators to avoid repetition when rendering “say” and other common verbs of saying in English, thus making use of a wider variety of verbs in the target texts with added connotations of meaning. There is also a tendency for reporting clauses with verb “say” to be omitted in the target texts. Likewise, Jesus (2004, 2008) examines an English-Brazilian Portuguese bidirectional parallel corpus of novels and finds that reporting clauses with the verb “say” in direct speech are frequently omitted in the translated texts into Portuguese.

Klaudy and Károly (2005) examine an English-Hungarian bidirectional parallel corpus of novels with the aim of exploring variability in the use of reporting verbs and its likely association with the phenomenon of explicitation in translation. They are particularly interested in investigating whether explicitation of meanings in one translation direction (English into Hungarian), as evidenced in Klaudy (2003), entails implicitation in the reverse direction (Hungarian into English). Their findings show that explicitation of meanings through the choice of reporting verbs takes place in both directions. Explicitation is carried out through the choice of verbs of saying with more specific meanings, indicating speakers’ tone, emphasis, state of mind, emotions, body movement, physical displacement, etc.

Similar findings are reported by Pipalovâ (2012), who analyses a Czech-English parallel and comparable corpus of novels. The starting point for her work is the assumption that Czech employs a wider variety of reporting verbs than English, which has an impact on translated texts in both directions. She finds that some of the reporting verbs in originals were turned into regular narrative in translated texts and others were omitted, especially when general verbs in English (“say”) were used, probably due to their being redundant. In a small number of cases, she finds addition of reporting verbs and narrative turned into direct speech.

Based on her comparison of non-translated text in Czech and English, Pipalovâ argues that, unlike the latter, where the reporting clause is considered redundant, in Czech, due to the richer variety of reporting verbs used by writers, the reporting clause itself plays a more integral role in narrative and character construction. Enhanced meanings pertaining to the speech act receptor, manner of speech, physical setting, emotional status, medium of interaction, among others, Pipalovâ concludes, contribute to a more engaging narrative for the prospective reader. They also contribute to the writer's controlling and guiding the reader's interpretation in that they offer more clues through reporting verbs as to how narrators act and feel when engaging in dialogic interaction. The implication is then that by introducing shifts in reporting verbs, translators also have an impact on reader's interpretation of translated narrative.

Wirf Naro (2012) arrives at a similar conclusion in her comparison of a German novel and its translation into French. For her, the greater variety of reporting verbs found in the translated text "deviates some of the reader's attention from the reported to the reporting"(116).

Vilinsky (2014) analyses a comparable corpus of English crime novels translated into Spanish and non-translated Spanish crime novels, a literary genre for which she argues dialogue is particularly relevant as a means to engage the reader in the crime or mystery resolution. She finds direct speech to be more frequent in novels translated from English into Spanish than in those written in Spanish, there being more reporting clauses in the translations which indicate aspects of the interaction. She finds more exclamatory utterances in the Spanish originals, which she posits as a form Spanish authors choose to qualify interactions. Regarding the variety of reporting verbs used in the translated texts, the results of her study confirm what previous researchers found: a wider variety of lexical verbs possibly due to the need to avoid repetition and to render "say" through different verbs in Spanish. This is suggested to have an impact on the reader as there is more explicitation of aspects of the speech act in the translated texts than the originals. As regards reporting clauses, the author finds that their presence is mainly motivated by the need to make it clear who is saying what, the Sayer being explicitly mentioned to further add to clarity. Martinez Vilinsky comments briefly on one of the drawbacks of adding variety in lexical verbs of saying in the target text. She explains that this may bring about a pattern that is radically different from that in the source text. This is the case when repetition of reporting verbs is stylistically motivated in the source text to construct a specific meaning in the narrative.

A contextual interpretation of translated text patterns is attempted by Alves (2006), who examines direct speech in two translations of Mark Twain's Huck Finn into Brazilian Portuguese, the first one published in 1934 and the second one in 1997. On the one hand, the two translated texts share features in common. Unlike the source text, which shows a high frequency of verb "say" in the present tense, the target texts evidence the use of an equivalent verb to "say" in simple past tense. On the other hand, Alves finds different patterns in the two target texts regarding the translation of reporting clauses and verbs. The 1934 translation tends towards the omission of reporting clauses when they involve the verb 'say', which adds to lesser narrator interference in the narrative; there is also a wider variety of lexical verbs to avoid repetition, verbs with an additional connotation such as animal noise being also included. The 1997 translation shows patterns more similar to those in the source text, with a higher frequency of reporting clauses with a neutral verb equivalent to “say” and less variety in other reporting verbs.

Drawing on Pagano (2001), Alves accounts for the patterns observed in the 1934 translation in terms of the publishing house and translator agenda of the time, which involved translating texts catering for an emerging audience, in the sense that readers were being introduced to reading mass market books in genres that were meant to arouse their interest for the reading experience.

Alves notes that the 1934 translation of Huck Finn is 24.5 % shorter than the original text by Mark Twain. The passages that were omitted in the translation correspond mostly to narrative passages, there being no dialogues omitted. This evidences, for Alves, a tendency for the translator to privilege the rendition of dialogic exchanges, which can be related to the purported idea of enhancing the prospective reader’s engagement with the text through the vivacity and dramaticity of dialogues.

The corpus analyzed in the study herein reported was compiled with a view to examining the claims made by Pagano (2001) and Alves (2006) through a corpus-based comparison of patterns in the translation of speech presentation in dialogic exchanges in novels first translated in Brazil in the 1930–1950s and later retranslated in the 2000s. This corpus configuration allowed us to both empirically verify claims regarding those translated texts and probe the so-called retranslation hypothesis briefly sketched in the following section.

4.6 Retranslation

Retranslation is a term used to name the practice of translating a text—usually an acclaimed literary or academic work—more than once into the same language. It generally, though not exclusively, implies the participation of different translators over a period of time. Antoine Berman (1990) first theorized retranslation in connection with the incomplete task of the translator in the sense that an original demands several translations until it can be most closely recreated in another language. Drawing on his reflections, the so-called “retranslation hypothesis” emerged, whereby retranslations tend to be closer to the original text than first translations, since they seek to improve what Berman posits as the intrinsic flaws of first attempts. Gambier (1994) pointed out the need for retranslations to be discussed along with other forms of rewriting, such as revised editions and adaptations, and research to be carried out regarding the motivations underlying the production of retranslations. This is addressed by Vanderschelden (2000, 4–6), who posits five motives for a work to be retranslated: (i) existing translations no longer meet expectations, regarding accuracy and/or adequacy and a revised edition is considered to be an inefficient solution to the problem; (ii) there is a new edition of the source text demanding target texts to be patterned upon it; (iii) the language and style of available translations need to be updated; (iv) there is a special need for a new target text due to market requirements; and (v) the new translation is presented as a new interpretation of the source text. As regards this last motive, Venuti (2003) suggests that it may be the translator's intent to produce a new translation of a previously translated source text.

Papoloski and Koskinen (2010) provide a thorough review of the main issues discussed in connection with retranslation and the most influential works on the topic. The authors argue that what research has shown is that the proximity relations between source text, first translations and retranslations are not necessarily a matter of chronology, but rather an issue having to do with the contextual parameters within which translations are performed. These apparently are closely related to commissioners and to translators themselves.

Papoloski and Koskinen do a survey of retranslations in Finland and confirm Vanderschelden (2000) in that the classification of a literary work as a "classic" makes it a likely candidate to be retranslated. However, the authors make a difference between reprints of translations and retranslations proper and ponder over the reasons why the latter would be opted for instead of the former. Their findings point to publisher profile as a factor in retranslations, more established publishing houses generally opting for reprints whereas emergent ones find in retranslations a way to grow, particularly in view that retranslations, Papoloski and Koskinen note, are more prone to be reviewed by the critic and thus gain publicity in the mass media.

In line with Papoloski and Koskinen, who stress the need for more inquiries into the topic of retranslation, the research herein reported seeks to address proximity relations in a corpus of retranslations as a case study where the contextual parameters conditioning the production of the translations can be related to sociocultural accounts of a historical period in particular, as explained in the following section.

4.7 The Golden Age of Translation

The corpus of retranslations herein analyzed is made up of three novels first translated in Brazil in the 1930–1950s and subsequently retranslated in the 2000s.

The 1930–1950 decades are known in Latin America as the Golden Age of Translation. Mexico city, Santiago de Chile, Rio de Janeiro, São Paulo, Porto Alegre and Buenos Aires were some of the main centers of translation and publishing during that period. Through their offices and printing presses, hundreds of authors from all historical periods and origins were put out into Spanish and Portuguese with a view to meeting the growing demand for reading and instruction of a new consumer class. Socioeconomic and political changes framed the ambience in which translators performed their task, the discourse of modernity reverberating in the expanding cities. Academic treatises and technical manuals, classic and contemporary novels, detective and hard boiled fiction, romances and comics filled the shelves of bookstores and newsstands. Publishing companies emerged and thrived, their catalogues advertising mostly translated texts. This translation boom was intricately framed by the ideology of modernity and instruction as social advancement in a period when industrial and labor market growth were pumping Latin American economies and contributing to the growth of urbanization and city development.

In perspective, the editorial projects put out by the publishing companies can be read as translation projects both in the narrow sense of the word, that is, as inter-linguistic translation, and in the broad sense of the term, that is, as cultural translation of art, knowledge and information. The specificity of this process together with its long lasting implications for signifying formations in society point to translation as a key cultural practice that informs negotiations inherent to cultural transference processes.

A consideration of the world and the Latin American contexts of the 1930–1950s account for the factors leading to the editorial and publishing *boom* in that period. The decades ranging from 1930 to 1950 saw a growing process of industrialisation and urbanisation. The labor market was expanding, allowing for a rise in the purchasing power of people who had hitherto had little access to consumer goods. A rise in people's personal income therefore meant a rise in consumption. Together with labor and economic changes, there were changes in the educational policies adopted by the government to improve basic education and literacy. These led to a rising number of literate people and hence of potential readers of books and magazines (Rivera 1980/86). Other sources of consumption involved new forms of mass media such as the cinema and the radio.

The relative isolation of Latin America after the 1929 world crisis and the new world map brought about by the Great and the Second World Wars motivated what critic Gerald Martin (1989) calls an introspective look on the part of Latin America. Trying to overcome the post war effects of the 1920s and the conflicts escalating to the Second World War confrontation, Latin America began to look inwards in order to reflect upon the representations of its own identities, musing upon concepts such as those of "Argentineness", "Brazilianness", "Mexicanness". This introspective turn in Latin American witnesses the birth and growth of cultural industries such as the book and the radio, together with the emergence of publishing houses and intellectual groups that seek to rethink the national within a new global context (Miceli 1979).

In the editorial scene, events such as the Spanish Civil War, which made innumerable publishing houses leave Spain and establish themselves in Argentina and Mexico (King 1989; Rama 1982), and the Second World War, which made it difficult for Latin American countries to import books (Hallewell 1982) and keep in touch with European metropolises, favored the publishing and translating effervescence of the 1930–1950s (Andrade 1978; Martins 1979; Miceli 1979; Hallewell 1982). Far from being detrimental to the publication of the works of local authors, translation contributed to the consolidation of the book market as a whole, helping towards the building up of a larger reading public (Miceli 1979) and thus fostering local authors' production (Hallewell 1982; Martins 1979). A sign of the growing editorial market at the time is the appearance of what turned out to be important publishing houses and printing presses, as was the case of Editora Globo, Companhia Editora Nacional, Martins and José Olympio in Brazil. Besides these major publishing houses, small houses and printing presses together with book clubs proliferated in the metropolises.

There was an institutional policy for the promotion of basic education and mass literacy. Together with these initiatives in education and reading skills development, there was a fertile literary scene, greatly favored by a task that significantly promoted most editorial enterprises: translation. Famous writers would work as translators for the emerging publishing houses, their literary careers showing a parallel development between their own writing and their translations. These writers' statements in autobiographies and memories reveal the significant role of their translation tasks in their own writings (Veríssimo 1973). Many publishing houses would publicize their translated texts making use of the name of their translators, often relevant figures in the national literary scenes, in covers, flaps and back-cover blurbs.

Translation involved finding a suitable way to render foreign authors into the vernacular languages of countries which had once been colonies of European metropolises and were striving to build a publishing market of their own. Thus, translating was performed into Argentine Spanish, Mexican Spanish and Brazilian Portuguese. The editorial and publishing activities in those decades share the debates' grounds with intense reflections about language and its national specificity. Retranslation of texts published in Portugal and Spain into the Latin American vernaculars was also carried out.

Still another factor leading to the increase in the editorial and translation activity of this period was the precarious state of copyright legislation at that time (Veríssimo 1973) which resulted in some publishers' not honoring copyright laws of the books they translated and not being legally penalized for that. Besides favoring a cheaper production of translated foreign books, the flexibility with which publishers dealt with copyright laws allowed for the multiple translations and editions of the same volume, thus offering the market different choices of price, binding, paper, size and, most interesting, of translator.

Unlike the learned readership belonging to a more affluent social class that had hitherto dominated book consumption in Latin America and who had always had access to national or foreign books, the new readers were not able to afford imported volumes, which required that in order for them to consume books, these had to be cheap and easy to purchase. Besides, the book had to fulfill not only a school or professional necessity but also a desire for leisure. Light fiction or entertainment books were therefore in demand and started to be read, particularly because of their high appeal, their cheap prices and their availability within readers' everyday surroundings (local bookstores, news-stands and other stores).

The great demand for books and magazines at that time led to the need to translate a wide range of reading materials ranging from novels, feuilletons and comics to academic treatises and textbooks. Translation also operated in spheres other than book publishing: newspapers published translated feuilleton novels, radio broadcasting companies aired translated soap opera scripts, film companies translated and adapted screenplays and most comics and cartoons in popular magazines were also translated (Haussen 1992; Rivera 1980/86).

Since the early 1930s the discourse of modernization and education as social advancement had been slowly infiltrating different social spheres in Brazil and publishers had begun to think of editorial projects that besides envisaging profits could lead to a popularization of reading and book selling in nations where people still had little access to educational and leisure activities. The idea was to turn books into mass-produced articles that could be marketed widely and in places where other consumer goods were offered for sale. In this sense, Brazilian publisher Monteiro Lobato's famous circular inviting store owners to sell books shows the desire to turn the book into an item of everyday consumption, something to be purchased and sold like other items such as soap, groceries and medicine. As an item, it was devised to be acquired by a new consumer, someone who had to be persuaded about the necessity and the potential advantages of buying that product. In fact, unlike the readers that had hitherto played a leading role in the publishing market, the new consumers did not possess inherited home libraries or consolidated reading skills; they did not master any foreign languages and did not have a substantial personal income either, which would have allowed them to consume imported volumes. They were actually a new reading public, an emerging middle class, partly composed by first generations of immigrant families, eager to seek educational and professional opportunities for personal advancement. Most significant, they were a new audience that was ready to enter the leisure consumer market, buying books not merely out of a necessity but also as a means of recreation (Lajolo and Zilberman 1996).

The need for translation arises thus out of a double need: on the one hand, the need to supply the industrial and education markets with treatises and technical manuals and school textbooks; on the other hand, the need to cater to, and at the same time stimulate, a growing demand for leisure products, such as popular genres of fiction.

An interesting fact about translation in the 1930–1950s is that the role of the translator was fulfilled by writers, editors and publishing house owners, such as, Érico Veríssimo and Monteiro Lobato in Brazil. Their statements and recollections show a particular concern with the fluency of their translations and rewritings, sometimes at the expense of having to introduce substantial modifications into the translated text (Veríssimo 1973; Lobato 1944).

The corpus analysed in the present study, as previously mentioned, is made up of three of the translations done and published during the Golden Age of Translation, later retranslated in the 2000s. Details about the corpus are provided in the following section.

4.8 Corpus

Three novels make up the corpus of this study: *Lord Jim* (Conrad 1971, 2002, 2005), *The thin man* (Hammett 1984, 1992, 2002) and *They shoot horses, don't they?* (McCoy 1982, 1995, 2000)

Joseph Conrad's *Lord Jim* was published as a serial in the British magazine Blackwood's between 1899 and 1900. It was first translated into Brazilian Portuguese by Mario Quintana and published under the title *Lorde Jim* (*Lord Jim*) in Brazil in 1939. The publishing house Editora Globo was at the time an emergent enterprise and *Lord Jim* was one of the titles of the book series Coleção Nobel (Nobel Collection), which aimed at presenting the Brazilian reader with translations of literary works acclaimed by critics. *Lord Jim* was subsequently retranslated in 1982 by Marcos Santarrita and published as *Lord Jim* by Editora Francisco Alves. A new retranslation came out in 2001 by Julieta Cupertino and was published as *Lord Jim, um romance* (*Lord Jim, a novel*) by Editora Revan. For the purposes of this study, the 1939 and the 2001 translations were selected.

Dashiell Hammett's *The thin man* was published in an American magazine, Redbook, in 1934. It was first translated into Brazilian Portuguese by Jose Bento Monteiro Lobato and published in Brazil in 1936 under the title *A ceia dos acusados* (*The supper of the suspects*), which was the title of the feature movie based on the novel released at that time. It was published within a book series called Série Negra (Noir series). It was retranslated by Rubens Figueiredo and published under the title of *O homem magro* (*The thin man*) by Companhia das Letras in 2002.

Horace McCoy's *They shoot horses, don't they?* was first published by the American publishers Simon and Schusters (1935). It was first translated into Brazilian Portuguese by Érico Veríssimo in 1947 and published under the title *Mas não se mata cavalo?* (*But don't they kill horses?)* by Editora Globo within the book series Coleção Tucano (Toucan Collection), a pocket edition series. It was retranslated by Renato Pompeu and published by Sá Editora in 2000 under the title *A noite dos desesperados* (*The night of the desperate*) borrowing the title of the feature movie based on McCoy's novel released in Brazil in 1969.

For the analysis carried out in this study, samples of the novels were used, manually selected following the criterion that they should include dialogic exchanges between characters in which verbal clauses were used for quoting and reporting. Further details about the procedures followed are provided in the following section.

4.9 Methodology

As previously mentioned, the choice of the corpus for this study was motivated by the wish to explore a translational corpus made up of source texts, first translations and retranslations, in order to examine patterns of use of direct speech in fictional dialogue in the source texts and their rendition in the target texts. An analysis of text equivalents found in the corpus in terms of formal correspondence and shifts was expected to yield an overall pattern for each translation and the results obtained to be interpreted in the light of statements made about tendencies in the translations analysed regarding translators' and publishers' agendas. The nature of the corpus would also allow for obtaining results that could be interpreted from the perspective of the so-called retranslation hypothesis and the study could contribute lending an empirical basis to that claim.

The corpus was compiled from fictional novels that met the following selection criteria:

- novels originally written in English by different authors
- novels first translated into Brazilian Portuguese and published in Brazil in the 1930–1950s and subsequently retranslated and published in Brazil in the 2000s
- novels translated and retranslated by different translators
- novels containing passages of dialogue in direct speech mode

The novels that fully met these requirements were *Lord Jim*, by Joseph Conrad, *The thin man*, by Dashiell Hammett and *They shoot horses, don't they?,* by Horace McCoy. Full bibliographical details of the source and the target texts can be found in the Reference section.

Each set of texts was labelled with reference to the surname of the original author (Con for Conrad; Ham for Hammett; and McCoy for McCoy) and an indication whether it was the source text (ST), the first translation or target text 1 (TT1) and its retranslation or target text 2 (TT2). Thus the labels used for the Conrad set were Con_ST, Con_TT2 and Con_TT2; for the Hammett set Ham_ST, Ham_TT1 and Ham_TT2; and for the McCoy set MCCoy_ST, McCoy_TT1 and MCCoy_TT2.

Ten samples of approximately 300 words containing dialogues were randomly selected and retrieved from the three source texts.

Equivalent passages to the source text samples were located in the target texts, retrieved and aligned with one another.

The aligned source and target text samples were segmented into clauses and all verbal clauses were copied and pasted to a sheet in a spreadsheet editor for manual annotation of categories pertaining to grammatical functions, lexical features, narratological features and equivalence relations, as defined in the theoretical framework developed. The categories annotated are shown in Table 4.25.

With the exception of the category "Lexical Verb", which was language specific, all other categories allowed for cross-linguistic comparisons.

For each category a variable "zero" was used to annotate instances of non-equivalence in any of the texts.

Annotation was carried out manually by two annotators and crosschecked for consistency by a third one.

Table 4.25 Categories annotated in the corpus samples

Grammatical functions	Order of saying	Activity_talking activity_targeting semiosis_non_projecting_no_verbiage semiosis_non_projecting_verbiage semiosis_projecting_quoting_imperating semiosis_projecting_quoting_indicating semiosis_projecting_reporting_imperating semiosis_projecting_reporting_indicating order.of.saying.ZERO
	Reception	non_reception reception reception.ZERO
	Semantic function	proposal proposition semantic.function.ZERO
	Type of verb	general_member verb_circumstantial_feature verb_manner_specifying_connotation verb_specific_demanding verb_specific_giving type.of.verb.ZERO
Lexis	Reporting verb	lexical_verb zero
Narratology	Narrative level	level_1 level_2 level_3 level_ZERO
Translation	Equivalence	correspondence shift zero

The programming language and software environment R (R Core Team 2014) was used to import the spreadsheets with the annotated samples and construct a dataframe. R scripts were developed for:

- Computation of distances between texts based on grammatical categories and generation of dendrograms;
- Computation of distances between texts based on equivalence categories and generation of dendrograms;
- Extraction of frequency counts for equivalence categories;
- Extraction of frequency counts for narrative level;
- Extraction of frequency counts for type of verb;
- Extraction of frequency counts for lexical verbs.

Hierarchical cluster analysis, an emergent technique in empirical translation studies (Ke 2012; Pagano 2014), was performed to measure distances and compute proximity between texts. Euclidean metrics were selected and the clustering method used was Ward's. All tables and figures were generated in R.

The next section presents the results obtained in our analysis.

4.10 Results

The first step carried out in our analysis consisted in comparing each set of STs, TT1s and TT2s through the technique of cluster analysis in order to see which relations of proximity could be established and to what extent these confirmed the retranslation hypothesis. The dataset used were the frequency counts for the categories annotated in each text referring specifically to grammatical functions.

The following dendrograms were obtained in R:

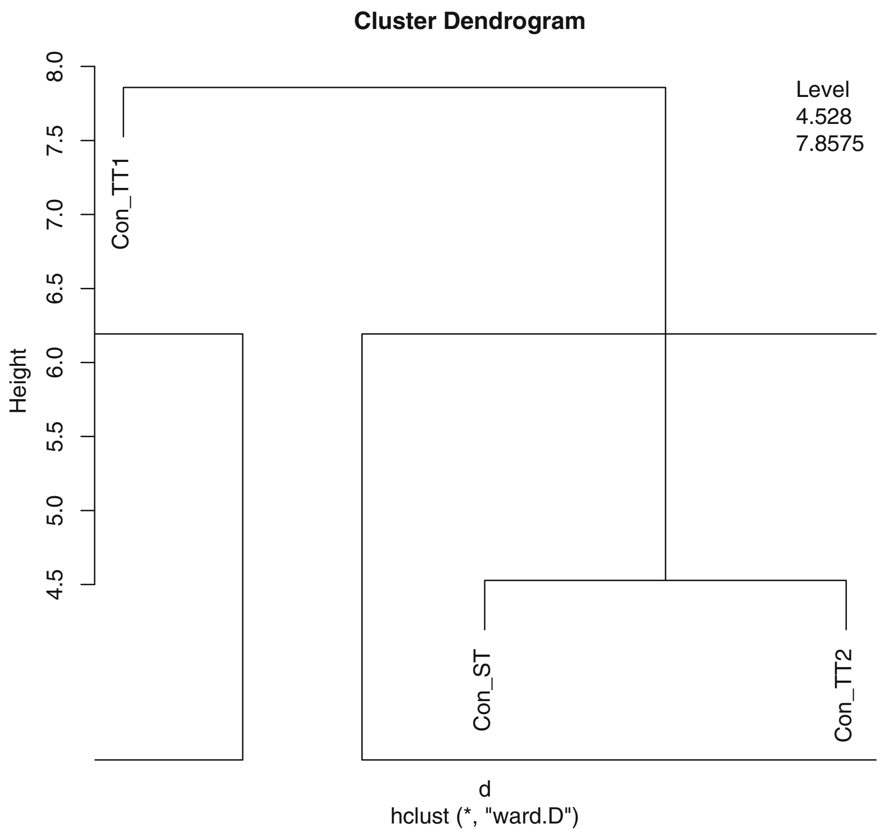


Fig. 4.1 Dendrogram showing hierarchical clustering of the Conrad set (Ward linkage) of Euclidean distances

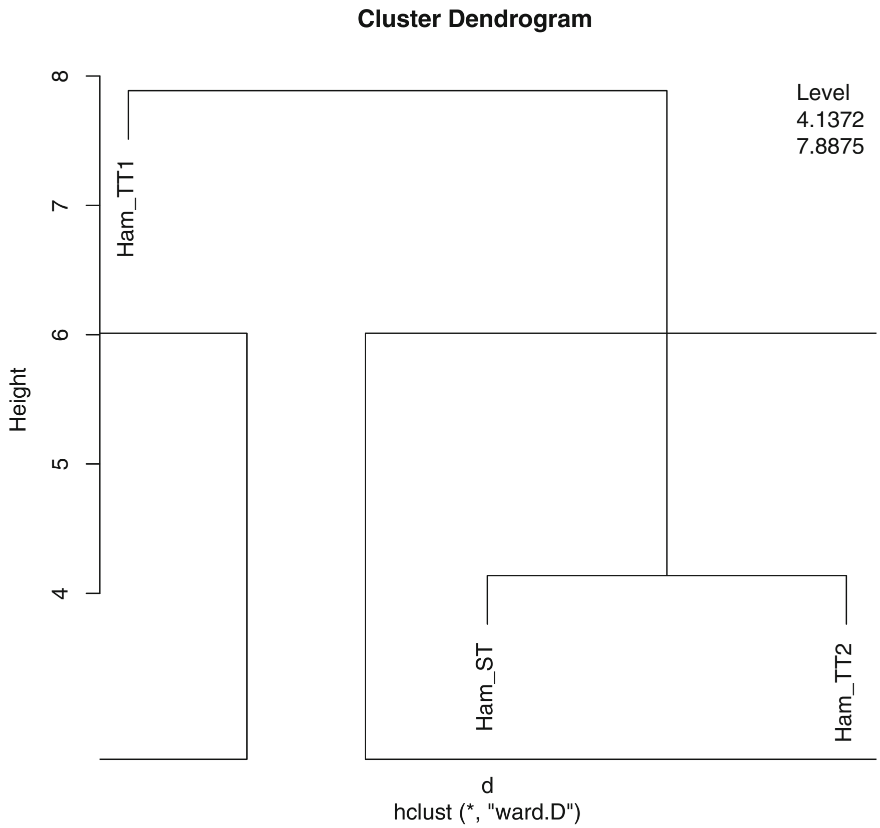


Fig. 4.2 Dendrogram showing hierarchical clustering of the Hammett set (Ward linkage) of Euclidean distances

The three dendrograms clearly cluster STs and TT2s in all three sample sets, TT1s being on another clade altogether. This means that the patterns in the use of verbal clauses relative to the functions annotated for each sample are more similar between STs and TT2s than between STs and TT1s or between TT1s and TT2s. The three sets of texts show a relatively similar level of fusion and the large jump for the two big clusters to fuse shows that the data is best represented by two clusters.

When the data from the three samples was combined in a single matrix, the following dendrogram was obtained:

The dendrogram in Fig. 4.4 is in agreement with the dendrograms obtained for each set of texts individually in Figs. 4.1, 4.2 and 4.3. Each ST clustered with its corresponding TT2, the three TT1s forming a rather distinct cluster, which shows that they are more similar to each other when compared to the other texts.

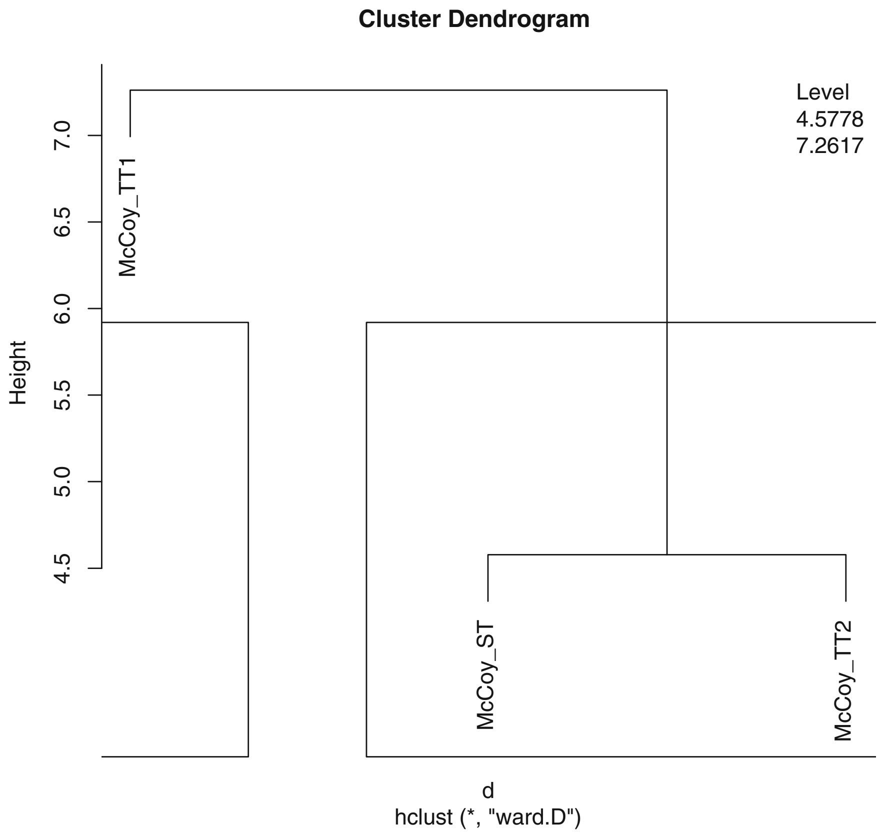


Fig. 4.3 Dendrogram showing hierarchical clustering of the McCoy set (Ward linkage) of Euclidean distances

The ST-TT2 cluster shows two subgroups, pointing to Conrad ST and TT2 as being less similar to the Hammett and McCoy STs and TT2s. This dissimilarity as will be seen later can be accounted for, among other variables, by the patterns in the use of lexical verbs of saying observed.

The proximity relations in the dendrogram shown in Fig. 4.4 confirm the retranslation hypothesis in that the first translation of a source text in a target culture tends to be further apart from the source text than subsequent translations of the same source text in that target culture.

A second step in our analysis was to cluster the translated texts in terms of equivalence patterns, based on frequency counts of occurrences annotated for formal correspondence, shift and zero equivalence. The following dendrogram was obtained in R (Fig. 4.5).

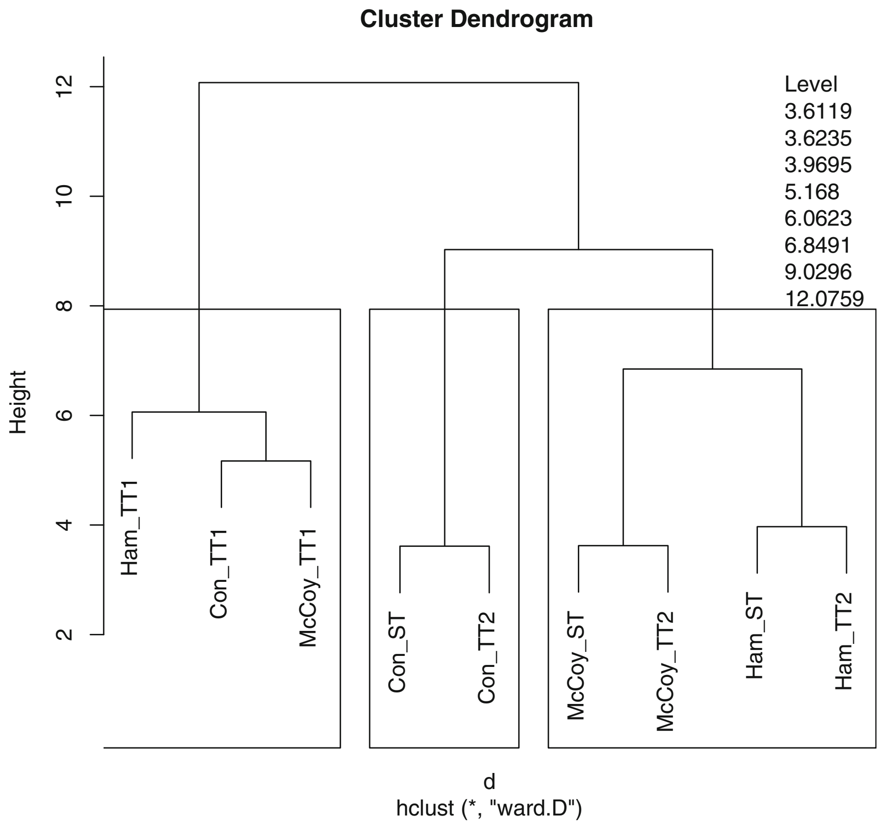


Fig. 4.4 Dendrogram showing hierarchical clustering of all texts based on grammatical categories (Ward linkage) of Euclidean distances

The results clearly show two main distinct clusters, made up by the three TT1s and the three TT2s, with a large jump until both are fused. This points to two groups of texts with relatively similar patterns of equivalence relations within each of them and very dissimilar between each other. This can be interpreted as two translational patterns: one by the translators in the 1930–1950s and another one by those in the 2000s. This also confirms another aspect of the retranslation hypothesis, namely the relative similarity of retranslations carried out within a particular historical period regardless of the different originals they render.

Among the TT1s, Hammet (Ham_TT1) stands out, fusing with Conrad (Con_TT1) and McCoy (McCoy_TT1) at a higher level. In order to verify this discrepancy, we can resort to relative frequency counts of equivalence and non-equivalence, the former comprising all verbal clauses in the STs that were rendered by verbal clauses in the TTs and the latter all verbal clauses in the STs that were omitted (not rendered) in the TTs or were added in the TTs but that were not in the STs.

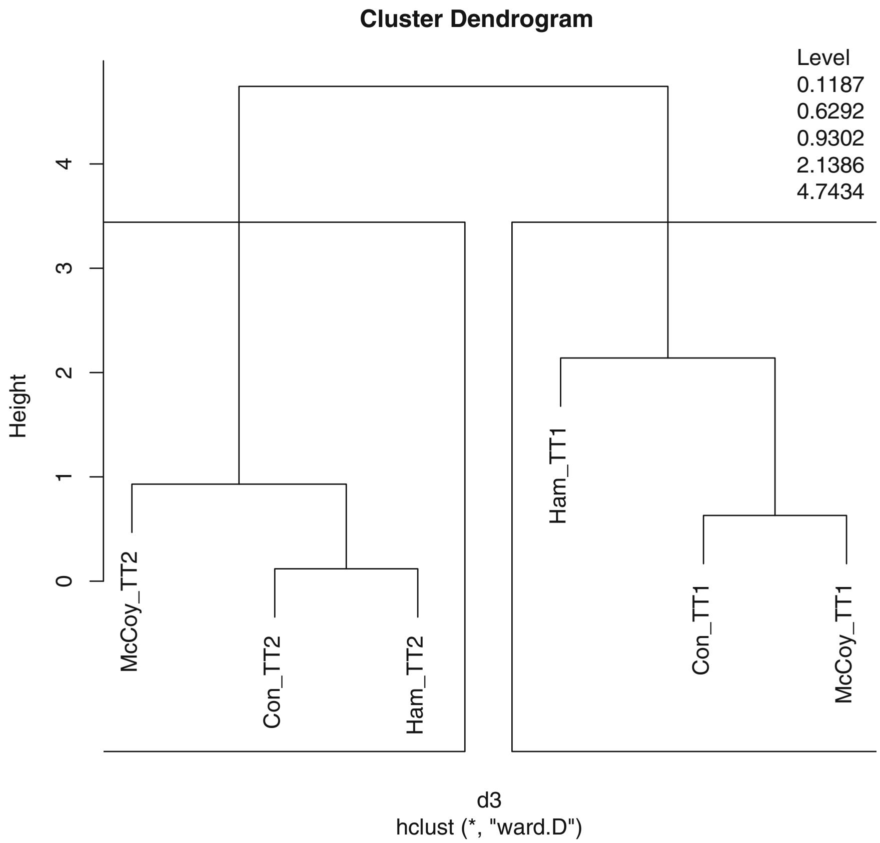


Fig. 4.5 Dendrogram showing hierarchical clustering of all texts based on equivalence categories (Ward linkage) of Euclidean distances

Relative frequency for equivalence categories counts are shown in Table 4.26. Table 4.26 shows that equivalence percentages are overall higher for TT2s than TT1s. However, Ham_TT1 shows a smaller percentage of equivalence and a bigger percentage of non-equivalence when compared to Con_TT1 and McCoy_TT1. This

Table 4.26 Percentage of equivalence in target texts (TT1s and TT2s)

	Equivalence (%)	Non-equivalence (%)	Total (%)
Con_TT1	68.85	31.15	100
Con_TT2	86.07	13.93	100
Ham_TT1	54.72	45.28	100
Ham_TT2	84.91	15.09	100
McCoy_TT1	75.00	25.00	100
McCoy_TT2	92.59	7.41	100

means that the number of verbal clauses in Ham_ST that were not rendered as verbal clauses in Ham_TT1 is not only higher than Ham_TT2, but also much higher than Con_TT1 and McCoy_TT1. The implication is then that translation equivalence patterns in Ham_TT1 point to a more idiosyncratic profile for the translator, Monteiro Lobato, which had been earlier pointed out by Alves (2006) in his comparison of Huck Finn's translation by Lobato (Hammett 1936) and its retranslation in 1997.

Equivalence percentages were calculated on the basis of frequency counts for the categories annotated pertaining to grammatical functions and equivalence. As explained in the Methodology, occurrences of non_equivalence had to be annotated for all other categories as being ZERO. Their mapping in the database with respect to any category of whatever type allows for explaining omission in the TTs and addition to STs.

The narratological category annotated—narrative level—is also instructive in a comparison of STs and TTs and TTS among each other. Table 4.27 shows the relative frequency for counts pertaining to the level in the narrative at which the dialogic exchange took place.

Table 4.27 Percentage of instances of verbal clauses at each narrative level

	Level_1 (%)	Level_2 (%)	Level_ZERO (%)	Level_3 (%)	Total (%)
Con_ST	76.03	16.53	6.61	0.00	100
Con_TT1	61.48	12.30	26.23	0.00	100
Con_TT2	76.23	14.75	9.02	0.00	100
Ham_ST	80.19	7.55	11.32	0.00	100
Ham_TT1	57.55	4.72	37.74	0.00	100
Ham_TT2	78.3	9.43	11.32	0.94	100
McCoy_ST	87.96	9.26	1.85	0.93	100
McCoy_TT1	66.67	8.33	24.07	0.93	100
McCoy_TT2	85.19	7.41	6.48	0.93	100

Most of the verbal clauses in STs, TT1s and TT2s, were located as operating at Level 1 in the narrative, that is, the level at which the narrator (or a character in the narrative) interacts with other characters. A lower number was located at Level 2, that is, the level at which characters quote or report conversations of other characters, and very few occurrences were located at Level 3, that is, they were verbal clauses in the speech reported by characters at Level 2, whose speech was in turn reported by characters or the narrator at Level 1. There were few shifts in narrative level and the main differences that can be observed when comparing the STs and TTs are in agreement with differences previously pointed out. These had to do with zero occurrences of verbal clauses (omission in the TTs and addition to STs), these being higher in TT1s than TT2s.

After grouping the texts on the basis of patterns of grammatical functions, the following step was to examine the only category annotated for lexis, which was the lexical verb used in each verbal clause. In order to do that, frequency lists of the reporting verbs used in each text were obtained.

Our analysis focused first on the frequency of reporting verbs in the STs that were rendered by reporting verbs in the TTs, i.e., equivalent verbs, and the frequency of verbs in the STs that were omitted (not rendered) in the TTs or verbs added in the TTs but that were not in the STs, i.e., non-equivalents. Tables 4.28, 4.29 and 4.30 show frequency counts for each set of texts.

Table 4.28 Frequency counts for equivalent and non-equivalent verbs in the Conrad set

	Con_ST	Con_TT1	Con_TT2
Equivalent verbs	116	90	112
Non-equivalent	6	32	10
Total	122	122	122

Table 4.29 Frequency counts for equivalent and non-equivalent verbs in the Hammett set

	Ham_ST	Ham_TT1	Ham_TT2
Equivalent verbs	94	66	94
Non-equivalent	12	40	12
Total	106	106	106

Table 4.28 above shows that TT2 follows closely ST pattern of use of reporting verb. TT1 shows a much higher number of instances of non-equivalence. Verbs that were not in ST but were used in TTs (additions) amount to 6.

Results for Hammett show an analogous pattern to the Conrad one, with figures for ST and TT2 being identical. TT1 shows a lower number of equivalent verbs and a higher one of non-equivalent ones.

The McCoy texts show patterns similar to Conrad's as can be seen in Table 4.30.

Table 4.30 Frequency counts for equivalent and non-equivalent verbs in the McCoy set

	McCoy_ST	McCoy_TT1	McCoy_TT2
Equivalent verbs	106	82	102
Non-equivalent	2	26	6
Total	108	108	108

Figures 4.6, 4.7 and 4.8 below show the proportion of equivalent and non-equivalent verbs for each set based on relative frequencies.

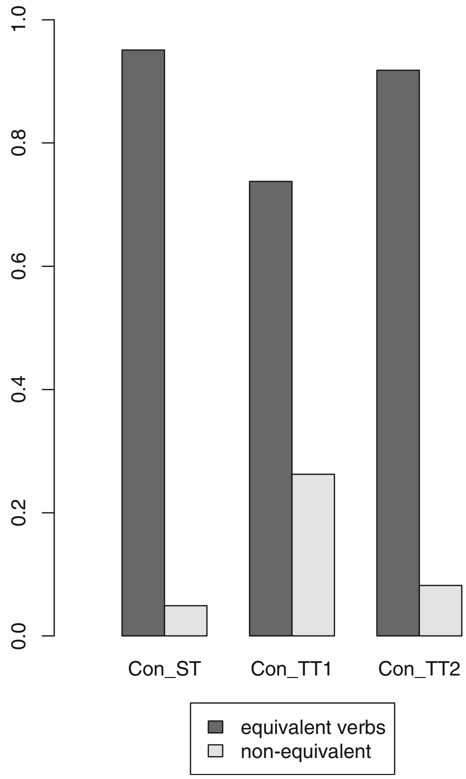


Fig. 4.6 Relative frequency of equivalent and non-equivalent verbs in the Conrad set

Relative frequencies show that equivalence in the use of reporting verbs is the main pattern, with Con_TT1 showing a higher proportion of non-equivalent verbs (omissions) than Con_TT2 and a small percentage of additions to Con_ST.

Hammett follows a analogous pattern, with Ham_TT1 showing a larger proportion o non-equivalent verbs than Ham_TT2 and than Con_TT1 above. More additions to Ham_ST can be observed when compared to those to Con_ST above. This distinct pattern in Hammett had already been pointed out in Fig. 4.4.

The McCoy set is more analogous to the Conrad set and it supports the general conclusion that equivalence prevails in all three sets, TT2s showing more similarity to STs than TT1s. TT1s evidence more omissions than TT2s.

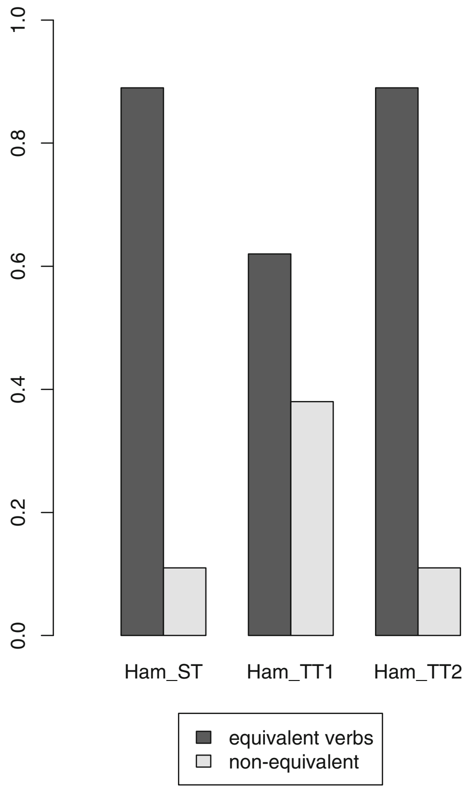


Fig. 4.7 Relative frequency of equivalent and non-equivalent verbs in the Hammett set

The second comparison centers on equivalent verbs to verify the proportion of general verbs—“say” in English and “dizer” (“say”) in Portuguese—and other verbs of saying. Frequency counts are displayed in Tables 4.31, 4.32 and 4.33.

Table 4.31 Frequency counts for occurrences of general verbs and other in the Conrad set

	Con_ST	Con_TT1	Con_TT2
Say/dizer	36	24	42
Other	80	66	70
Total	116	90	112

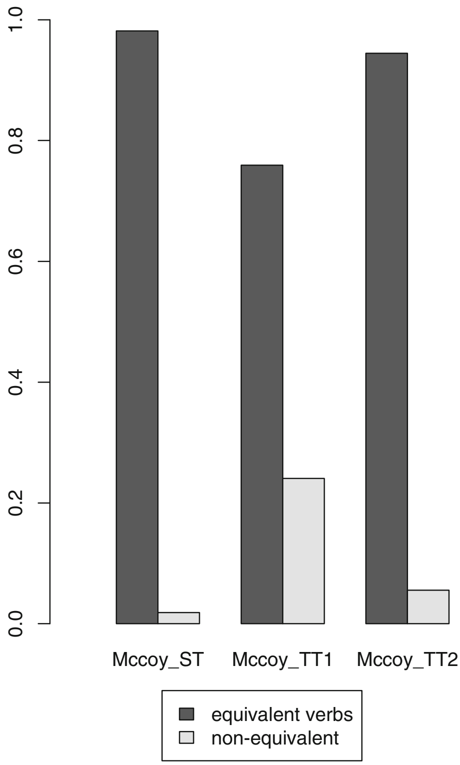
Table 4.32 Frequency counts for occurrences of general verbs and other in the Hammett set

	Ham_ST	Ham_TT1	Ham_TT2
Say/dizer	59	14	33
Other	35	52	61
Total	94	66	94

Table 4.33 Frequency counts for occurrences of general verbs and other in the McCoy set

	McCoy_ST	McCoy_TT1	McCoy_TT2
Say/dizer	76	32	42
Other	30	50	60
Total	106	82	102

Fig. 4.8 Relative frequency of equivalent and non-equivalent verbs in the McCoy set

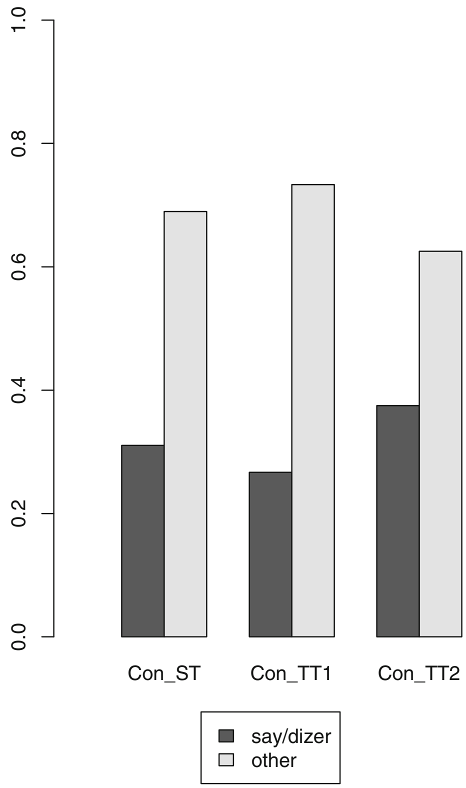


The relative proportion of occurrences of "say"/"dizer" in relation to other verbs can be visualized in Figs. 4.9, 4.10 and 4.11.

The particular profile of the Conrad set had been pointed out in Fig. 4.4. Here we can clearly see that, unlike reports in the literature stating that "say" is the norm and most frequent lexical verb in speech presentation, Con_ST has a higher percentage of verbs other than "say" and this is somehow reflected in Con_TT1 and Con_TT2. Con_TT1 still shows a distinct pattern as it has a slightly higher percentage of verbs other than "say" when compared with Con_ST and Con_TT2.

Figures 4.10 and 4.11 below present the Hammett and McCoy data, both conforming to the pattern of "say" as a general verb frequently used in English.

Fig. 4.9 Relative frequency of verbs "say"/"dizer" and other verbs in the Conrad set



In Fig. 4.10, "say" appears in a high proportion in Ham_ST whereas this sharply decreases in Ham_TT1 and less markedly in Ham_TT2.

In Fig. 4.11, the McCoy set follows an analogous tendency, even though unlike Hammett, the difference in the use of verbs other than say in McCoy_TT1 and McCoy_TT2 is very thin.

The clear implication of the above results is that the three source texts in our corpus show heterogeneity in the patterns of use of reporting verbs. The source texts by McCoy and Hammett show little diversity of verbs other than "say", which conforms to the pattern pointed out in the literature of "say" being the general member of the class of verbs serving as process in verbal clauses used to quote (Halliday and Matthiessen 2014; Leech and Short 1981; Semino and Short 2004). In contrast, the source text by Conrad has a far wider variety of verbs other than "say" and the frequency of occurrence of "say" is lower than expected. This may be related to literary genre, since both the McCoy and the Hammett novels can be classified as crime novels (Glover 2003), whereas Conrad's is usually classified as

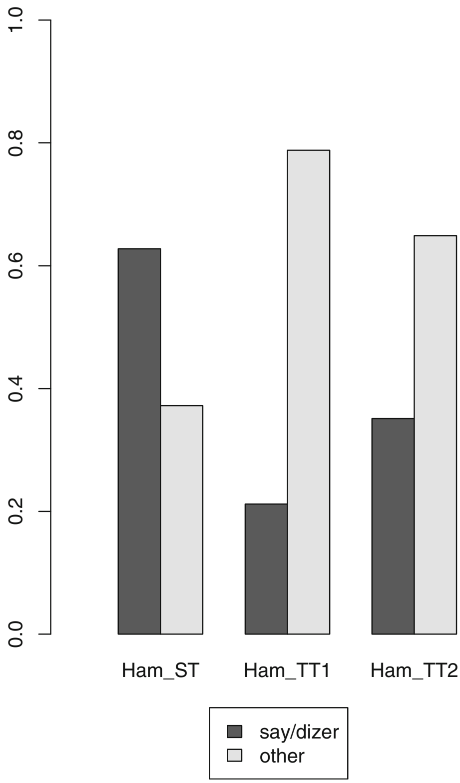


Fig. 4.10 Relative frequency of verbs "say"/"dizer" and other verbs in the Hammett set

an adventure novel with an experimentation in the genre of Bildungsroman or 'novel of education' (Stape 1996).

A close inspection of those verbs other than "say" is revealing in terms of variety of different verbs. This is computed by counting the number of different verbs regardless of their frequency of occurrence. The results are displayed in Tables 4.34, 4.35 and 4.36.

Table 4.34 Number of different lexical verbs other than "say"/"dizer" in the Conrad set

	Con_ST	Con_TT1	Con_TT2
Different lexical verbs Other than say/dizer	42	35	34

Table 4.35 Number of different lexical verbs other than "say"/"dizer" in the Hammett set

	Ham_ST	Ham_TT1	Ham_TT2
Different lexical verbs Other than say/dizer	14	26	16

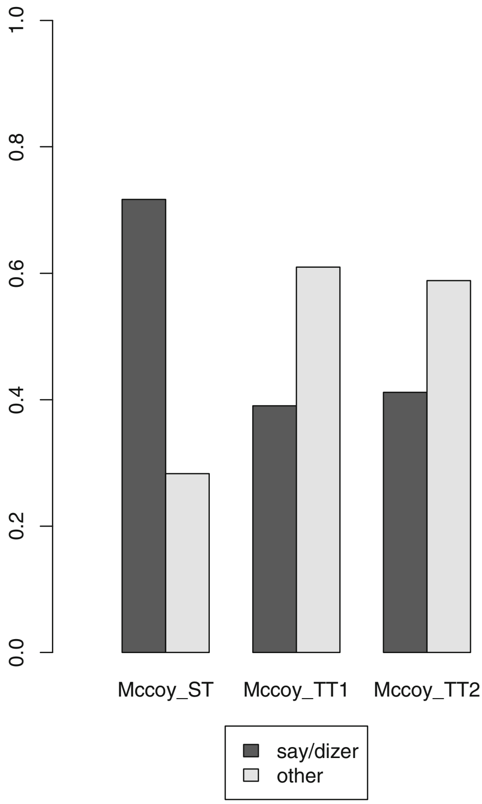


Fig. 4.11 Relative frequency of verbs “say”/“dizer” and other verbs in the McCoy set

Table 4.36 Number of different lexical verbs other than “say”/“dizer” in the McCoy set

	McCoy_ST	McCoy_TT1	McCoy_TT2
Different lexical verbs Other than say/dizer	10	24	20

As can be seen, Con_ST has a wide variety of reporting verbs other than “say”, which decreases in 12 and 15 % in Con_TT1 and Con_TT2 respectively. This is in sharp contrast with the figures for Hammet and McCoy:

Ham_ST’s variety of verbs increases by 86 % in Ham_TT1 and 14 % in Ham-TT2, showing close proximity between Con_ST and Con_TT2.

More noticeably, McCoy_ST’s variety of verbs increases by 140 and 100 % in McCoy_TT1 and McCoyTT2 respectively. The numbers also confirm close proximity between McCoy_ST and McCoy_TT2, even though the difference between McCoy_TT1 and McCoy_TT2 is much less marked than in the case of Hammett.

The different tendencies in the three text sets can be visualized in Fig. 4.12.

Fig. 4.12 Number of different lexical verbs other than “say”/“dizer”

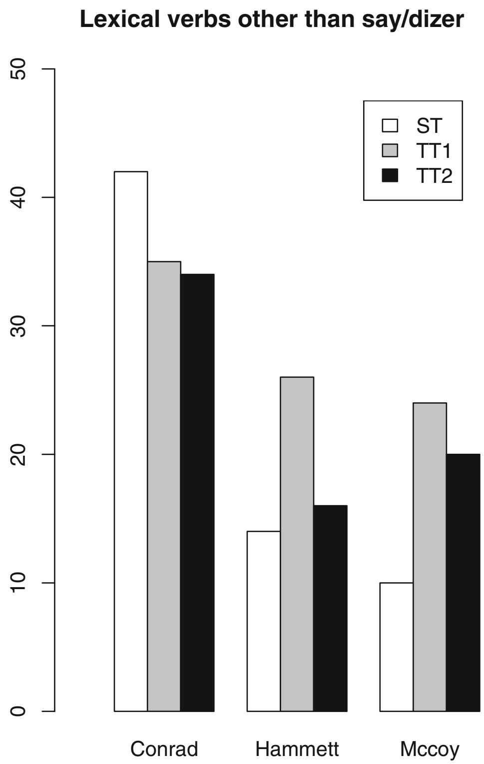


Figure 4.12 reveals the richer variety of lexical verbs in Conrad’s source text, which does seem to have an impact on the target texts. Nonetheless, however different the Conrad set might be, the pattern in proximity relations between STs, TT1s and TT2s is still kept, in that TT2s are closer to STs.

At this stage in the analysis, a question emerges: what makes TT1s different from TT2s and what kind of shifts can be identified? The answer to this question naturally leads us to asking how we can best account for the patterns found and within what contextual environment equivalence was established.

A first step was to look at the particular occurrences of non-equivalent verbs in the source text that are realized in the target texts and vice versa, realized verbs in the source text that have non-equivalents in the target texts.

With regard to non-equivalents in the ST, that is, meanings rendered in the TTS that were not realized in the ST, Table 4.37 yields the following data.

There are seven occurrences of lexical verbs of saying realized in TTs that are not realized in the ST. As can be seen, Con_TT1 tends to realize fewer of those meanings through lexical verbs than Con_TT2. In both TTs, the verbs of saying chosen to realize zero occurrences in the source text are general verbs (“dizer”/

Table 4.37 Non-equivalent lexical verbs in Conrad source text

Row.names	Con_ST	Con_TT1	Con_TT2
5	zero	zero	dizer
14	zero	gritar	zero
17	zero	afirmar	zero
36	zero	zero	falar
38	zero	zero	perguntar
50	zero	zero	comentar

Table 4.38 Non-equivalent lexical verbs in Conrad Target Text 1

Row.names	Con_ST	Con_TT1	Con_TT2
5	zero	zero	dizer
11	speak	zero	zero
27	say	zero	dizer
28	say	zero	dizer
29	say	zero	zero
33	tell	zero	confessar
34	say	zero	dizer
36	zero	zero	falar
37	flatter	zero	zero
38	zero	zero	perguntar
39	go on	zero	prosseguir
43	say	zero	zero
48	murmur	zero	murmurar
50	zero	zero	comentar
56	snarl	zero	rosnar
58	repeat	zero	repetir
87	say	zero	dizer
89	add	zero	acrescentar
90	whisper	zero	sussurrar
93	say	zero	dizer
95	stammer	zero	gaguejar
97	say	zero	dizer
98	admit	zero	admitir
99	say	zero	dizer
100	mumble	zero	resmungar
102	say	zero	perguntar
103	say	zero	dizer
118	call	zero	zero
119	say	zero	dizer
120	talk	zero	falar
121	talk	zero	falar
122	tell	zero	recomendar

"say"), verbs specific to speech function ("perguntar"/"ask"), verbs specifying manner ("gritar"/"shout") and verbs adding a circumstantial feature ("comentar"/ "comment"). Table 4.38 shows non-equivalent lexical verbs in Con_TT1.

We can clearly see that Con_TT1 tends not to realize many of the lexical verbs in Con_ST, particularly when those are the general verb "say" and verbs of saying as activity such as "speak" and "talk". Some of the verbs specifying manner, such as "stammer" and "mumble" were not realized in Con_TT1, since not only reporting clauses but reported ones and whole stretches of narrative for that matter were omitted in Con_TT1.

A similar trend, although in much lesser proportion, can be observed in non-equivalent lexical verbs in Con_TT2, as Table 4.39 shows.

Table 4.39 Non-equivalent lexical verbs in Conrad Target Text 2

Row.names	Con_ST	Con_TT1	Con_TT2
11	speak	zero	zero
14	zero	gritar	zero
17	zero	afirmar	zero
29	say	zero	zero
37	flatter	zero	zero
43	say	zero	zero
62	say	explicar	zero
76	tell	dizer	zero
80	tell	dizer	zero
118	call	zero	zero

In the case of the Hammett samples, unlike in the Conrad samples, we can observe that unrealized meanings in the Source Text tend to be realized by verbs of saying more frequently in Target Text 1 than in Target Text 2, which suggests more proximity between Source and Target Text 2, Target Text 1 being further distanced from the Source Text, which confirms the proximity relations obtained in the dendrograms.

Table 4.40 shows that verbs realized in Ham_TT1 that were not realized in Ham_ST are mostly verbs with a circumstantial feature signalling stages in turn taking, such as "indagar" ("inquire"), "perguntar" ("ask") and "responder" ("reply"). In one case only, both target texts add a verb of saying not realized in Ham_ST, which may indicate that both translators felt the need to do that. Interestingly, as Table 4.41 shows, this is a case of a verb of saying implicit in a reported clause in the original text:

The quoted clause "Try his lawyer" has en elliptical verb of saying as activity and could be worded as "try to talk to his lawyer". Both TTs do away with that ellipsis making use of two lexical verbs of saying: "indagar" ("query") and "falar" ("talk"), thus rendering meanings in a more explicit way.

Table 4.40 Non-equivalent lexical verbs in Ham_ST

Row.names	Ham_ST	Ham_TT1	Ham_TT2
12	zero	indagar	falar
15	zero	zero	falar
22	zero	indagar	zero
37	zero	perguntar	zero
43	zero	dizer	zero
45	zero	indagar	zero
72	zero	responder	zero
77	zero	murmurar	zero
79	zero	contestar	zero
85	zero	zero	falar
105	zero	dizer	zero

Table 4.41 Verb of saying implicit in Ham_ST and explicitated in Ham_TT1 and Ham_TT2

Ham_ST	"Try his lawyer," I suggested
Ham_TT1 *gloss*	- Por que não indaga do seu advogado?—sugeri *Why don't you query his lawyer?—I suggested*
Ham_TT2 *gloss*	- Tente falar com o advogado dele—sugeri *"Try to talk with his lawyer"—I suggested*

Table 4.42 shows non-equivalents in Ham_TT1.

The number of non-equivalent verbs in Ham_TT1 is relatively high and the majority of them do not render "say" in Ham_ST. Curiously, verbs that were added in Ham_TT1 but were not realized in Ham-ST (Table), like "ask", are now omitted in Ham_TT1. This apparent contradiction points to the need to complement frequency counts with actual discourse analysis to investigate the role of reporting verbs in narrative progression.

The example in Table 4.43 shows how an initiating turn in an exchange is recreated in Ham_ST, Ham_TT1 and Ham_TT2, reporting verbs underlined.

Ham_ST constructs an exchange patterned upon the conventions in English, whereby "say" as a general verb of saying is also used for replies. Ham_TT1 does not make use of a reporting clause for the initial interrogative clause, which the reader can infer to be a question through punctuation, but still makes use of a verb of saying with a circumstantial feature ("reply") closing the initial exchange. Ham_TT2, unlike Ham_TT1, initiates the exchange with a verb signalling speech function in an initial turn ("perguntar"/"ask") and closes it with a verb of saying with a circumstantial feature ("responder"/"reply"). The implication here is that the choice of whether or not to use a reporting verb and which one to use may have more to do with the unfolding of text in an exchange than with patterns of more frequent verbs in each language system and type of situation.

Non-equivalents in Ham_TT2 as Table 4.44 shows are fewer than in Ham_TT1 and are mostly in contrast with additions in Ham_TT1 which were not realized in Ham_ST.

Table 4.42 Non-equivalent lexical verbs in Ham_TT1

	Ham_ST	Ham_TT1	Ham_TT2
1	ask	zero	perguntar
3	say	zero	responder
9	say	zero	dizer
15	zero	zero	falar
19	say	zero	falar
24	break off	zero	zero
26	say	zero	explicar
27	say	zero	falar
28	say	zero	dizer
29	ask	zero	perguntar
30	say	zero	ordenar
32	say	zero	responder
33	ask	zero	zero
36	say	zero	dizer
38	say	zero	dizer
39	say	zero	dizer
40	say	zero	responder
46	say	zero	contar
47	tell	zero	dizer
49	say	zero	responder
53	say	zero	responder
55	talk	zero	falar
56	explained	zero	explicar
62	tell	zero	dizer
63	tell	zero	dizer
65	say	zero	falar
66	say	zero	dizer
67	zero	zero	falar
68	say	zero	responder
69	grumble	zero	resmungar
70	say	zero	dizer
78	say	zero	responder
80	say	zero	dizer
81	ask	zero	perguntar
84	say	zero	dizer
85	zero	zero	falar
86	say	zero	responder
96	say	zero	responder
101	whisper	zero	sussurrar
102	say	zero	responder

Table 4.43 Dialogic exchange in Ham_ST and its rendition in Ham_TT1 and Ham_TT2

Ham_ST	'Aren't you Nick Charles?' she asked I said 'Yes'
Ham_TT1	- Nick Charles?
gloss	*Nick Charles?*
	- Sim—respondi.
gloss	*-Yes—I replied*
Ham_TT2	- Você não é Nick Charles?—perguntou ela
gloss	*- Aren't you Nick Charles?—asked she*
	- Sou—respondi
gloss	*– I am—I replied*

Table 4.44 Non-equivalent lexical verbs in Ham_TT2

	Ham_ST	Ham_TT1	Ham_TT2
22	zero	indagar	zero
24	break off	zero	zero
25	ask	indagar	zero
33	ask	zero	zero
37	zero	perguntar	zero
42	assure	continuar	zero
43	zero	dizer	zero
45	zero	indagar	zero
72	zero	responder	zero
77	zero	murmurar	zero
79	zero	contestar	zero
105	zero	dizer	zero

Table 4.45 Non-equivalent lexical verbs in McCoy_ST

	McCoy_ST	McCoy_TT1	McCoy_TT2
42	zero	dizer	dizer
68	zero	zero	anunciar

In the case of the McCoy set, non-equivalents in MCCoy_ST are only two and they represent additions of "dizer" ("say") and "anunciar" ("announce") made in McCoy_TT1 and McCOy_TT2 (Table 4.45).

The number of non-equivalents increases considerably in McCoy_TT1, mostly as omissions of "dizer" ("say") (Table 4.46).

The number of non-equivalents noticeably decreases in MCCoy_TT2 and has to do with omissions mostly of "say" (Table 4.47).

Again it is instructive to have a look at some of the instances of omissions from the perspective of narrative unfolding. Table 4.48 brings an example, reporting verbs underlined and gloss underneath.

The example shows that the detective is talking simultaneously with Mario, a.k.a Giuseppe, and with the narrator. Whereas in MCCoy_ST "say" is used three times

Table 4.46 Non-equivalent lexical verbs in McCoy_TT1

	McCoy_ST	McCoy_TT1	McCoy_TT2
2	ask	zero	perguntar
4	say	zero	dizer
5	talk	zero	falar
8	say	zero	pedir
11	say	zero	dizer
12	say	zero	falar
14	ask	zero	perguntar
15	say	zero	dizer
16	say	zero	zero
17	say	zero	dizer
19	say	zero	responder
29	say	zero	dizer
40	say	zero	zero
48	ask	zero	zero
49	say	zero	dizer
57	ask	zero	perguntar
68	zero	zero	anunciar
69	say	zero	dizer
72	say	zero	responder
73	say	zero	dizer
74	say	zero	dizer
79	say	zero	dizer
93	say	zero	perguntar
94	say	zero	zero
95	say	zero	responder
102	say	zero	dizer

Table 4.47 Non-equivalent lexical verbs in McCoy_TT2

	McCoy_ST	McCoy_TT1	McCoy_TT2
16	say	zero	zero
40	say	zero	zero
48	ask	zero	zero
94	say	zero	zero
97	say	dizer	zero
101	tell	dizer	zero

for those exchanges, MCCoy_TT1 makes use of a verb explicitly signalling speech function ("sugerir"/"suggest") to introduce the detective's command, which is resumed in the second turn with a verb shaving an additional circumstantial feature ("continuous"/"go on") and finally omits the verbal cause "he said to Mario", avoiding the double identity of Mario/Giuseppe. McCoy_TT2 renders the three reporting verbs in MCCoy_ST through verbs that shift the commands uttered by the

Table 4.48 Reporting verbs in the flow of discourse

McCoy_ST	'Never mind your partner, Giuseppe,' the second detective said 'Hey son,' he said to me, 'you tell Giuseppe's partner good-bye for him. Come on Giuseppe,' he said to Mario
McCoy_TT1	– Deixe o seu par, Giuseppe—sugeriu o segundo detetive – *Leave your partner, Giuseppe—suggested the second detective* – Você aí, meu filho—continuou ele, dirigindo-se a mim –, diga adeus pelo Giuseppe à companheira dele. Vamos, Giuseppe – *You there, son—he went on, addressing me—say goodbye to Giuseppe's partner in his name. Come on, Giuseppe*
McCoy_tt2	"Não se incomode com a parceira, Giuseppe", falou o segundo policial *Don't worry about your partner, Giuseppe, spoke the second policeman* "Ei, filho", me chamou, "você se despeça por Giuseppe da parceira dele. Venha, Giuseppe", disse ele a Mario *Hey, son, he summoned me, you say goodbye to Giuseppe's partner in his name. Come, Giuseppe, said he to Mario*

detective to propositions being quoted in the dialogue. Overall, the authoritarian voice of the detective in McCoy_ST and MCCoy_TT1 becomes less so in McCOY_TT2. Shifts of this kind are worth investigating in terms of their cumulating effect in the text, a perspective which exceeds the scope of the present study.

A last stage in our analysis consisted in observing the main types of shifts in each set of texts in order to verify to what extent they could point to patterns in TT1s and TT2s. Table 4.49 shows the results for the Conrad set.

Table 4.49 Major types of shift from Con_ST to Con_TT1 and Con_TT2

SHIFTS	Con_TT1	Con_TT2
non_reception TO reception	6	1
verb_specific_giving TO general_member	4	4
verb_manner_specifying_connotation TO verb_circumstantial_feature	3	1
proposal TO proposition	1	2
reception TO non_reception	1	2
verb_manner_specifying_connotation TO general_member	1	1
verb_specific_giving TO verb_manner_specifying_connotation	1	1

The shift from non reception, i.e., no Receiver for the verbal process realized by the verb of saying, to reception is more frequent in Con_TT1 than in Con_TT2. This has to do with explicitating who is being addressed in an exchange, as illustrated by the example in Table 4.50.

Con_ST makes use of three reporting verbs, one of them ("ask") signalling speech function (demanding a proposition) and the other two ("say") quoting propositions. Con_TT1 brings two shifts: a general verb "say" to a verb with an

Table 4.50 Shift in the system of reception/non-reception

Con_ST	"What can you see?" he asked "Nothing," I said, feeling awfully ashamed of myself "Just so," he said
Con_TT1	Que é que o senhor vê? perguntou êle. – *What is it that you see? asked he* Nada respondi, envergonhado de mim próprio *Nothing, I replied, ashamed of myself* Naturalmente! disse-me *Naturally! he said to me*
Con_TT2	"O que o senhor está vendo?"—perguntou ele *What can you see?—asked he* "Nada", disse eu, sentindo-me terrivelmente envergonhado de mim mesmo. *Nothing, said I, feeling terribly ashamed of myself* "Isso mesmo", disse ele, *That's it, said he*

additional circumstantial feature ("responder"/"reply") and introduces a Receiver ("me"/"me") explicitating that the narrator was the addressee of the interlocutor's remark "Just so". Con_TT2 does not show any shift from the patterns in Con_ST.

In the Hammett set, the major types of shifts are shown in Table 4.51.

Table 4.51 Major types of shift from Ham_ST to Ham_TT1 and Ham_TT2

Shifts	Ham_TT1	Ham_TT2
general_member TO verb_manner_specifying_connotation	11	1
general_member TO verb_circumstantial_feature	10	6
general_member TO verb_specific_giving	2	3
verb_specific_demanding TO general_member	1	2

The largest number of shifts, and particularly in Ham_TT1, has to do with realizing verb "say" in Ham_ST as a verb specifying manner or with an additional circumstantial feature, as the examples in Table 4.52 shows.

Table 4.52 Shift in type of verb in the Hammett set

Ham_ST	'FIRST thing,' Guild said as we left his office, 'we'll go see Mr Nunheim
Ham_TT1	- A primeira coisa que temos a fazer é procurar o Sr. Nunheim—advertiu Guild
Ham_TT2	Em primeiro lugar—disse Guild, ao sair do seu gabinete-,Viemos ver o senhor Nunheim.

The general verb "say" in Ham_ST is rendered as "advertir" ("warn") in Ham_TT1, a verb with an additional circumstantial feature, signalling an undesirable result if something is not done as suggested. This meaning seems to render an implicit command in Ham_ST, which could be worded as "Please listen to me and follow my orders". In Ham_TT2, "dizer" ("say") is used, but interestingly it is used to quote an implicit proposal, which could be worded as "Remember that our target is to see Mr Nunheim". Once again, looking at how verbs are used in the development of an exchange can throw more light into understanding why shifts take place.

Finally, Table 4.53 brings the major types of shift in the MCCoy set.

Table 4.53 Major types of shift from McCoy_ST to McCoy_TT1 and McCoy_TT2

Shifts	McCoy_TT1	McCoy_TT2
general_member TO verb_circumstantial_feature	11	9
general_member TO verb_specific_giving	8	11
general_member TO verb_specific_demanding	6	5
general_member TO verb_manner_specifying_connotation	3	3
verb_specific_giving TO general_member	2	1
verb_specific_giving TO verb_circumstantial_feature	1	1

Similarly to the Hammett set, most of the shifts in the McCoy set have to do with type of reporting verb, their number being equally expressive in McCoy_TT1 than in McCoy_TT2. The example in Table 4.54 shows a shift from the general verb "say" to a verb with an additional circumstantial feature "reply" in both TTs.

Table 4.54 Shift of type of verb in the McCoy set

McCoy_ST	'What happened?' Socks asked her 'He accused me of cheating–' she said
McCoy_TT1	– Que foi que aconteceu?—perguntou-lhe Socks *What is it that happened? asked her Socks* – Pedro me acusou de estar enganando-o—respondeu ela *Pedro accused me of cheating him—replied she*
McCoy_TT2	"O que aconteceu?", perguntou Socks *What happened? asked Socks* "Ele me acusou de enganá-lo", respondeu ela *He accused me of cheating him—replied she*

The example above illustrates a shift which can be accounted for by looking at the development of the dialogic exchange: "say" used to signal a response to an initiating move in which a question is posed is rendered in the TTs by "reply", clearly explicitating the closing move in that exchange.

The results hitherto presented will be summarized in the coming section and discussed in the light of previous findings in the literature.

4.11 Discussion

The results obtained through hierarchical clustering of texts based on the grammatical categories annotated showed a distinct proximity relation between STs and TT2s for all three sets. TT1s form a separate group away from STs and TT2s. The same proximity pattern was observed based on equivalence categories annotated for each text.

These results support the retranslation hypothesis as posited by Berman (1990) and suggest that the first translations carried out in the 1930–1950s were actually a first step in the introduction of those originals into the target culture. The number of non-equivalents (omissions and additions) and the type of shifts observed may be accounted for at the contextual level (context of culture).

The results also support reassessments of the retranslation hypothesis as presented by Paloposki and Koskinen (2010), in that retranslations are the outcome of publishing and translation projects within particular historical conditions. In the particular case of the study herein reported, first translations emerged as having clear distinct patterns in the use of reporting verbs. The three first translations (TT1s) analysed showed a similarity profile despite being renditions of different source texts by different translators. An aspect in common to all three was their being produced by young publishing houses in a particular historical period, as described in section Translation Booms, and the fact that the three translators were writers themselves, two of them owning publishing houses, with relative autonomy as to how to direct the translation process.

Unlike first translations, the retranslations analysed (TT2s) were published in a different historical period, when the publishing market in Brazil was consolidated and readers had a different profile from those envisaged in the 1930–1950s. The three retranslations were done by professional translators working under the commission of publishing houses.

The conditions under which first translations and retranslations were carried out can be theorized within a comprehensive theory of language through the notion of meta-context, as proposed by Matthiessen (2001). This allows for explaining equivalence and shifts from the most local contextual environment of the clause to the most global one of the context of culture.

In this respect, the patterns observed in the use of lexical verbs in source and target texts are in agreement with the descriptions available in linguistic and stylistic accounts (Halliday and Matthiessen 2014; Leech and Short 1981; Semino and Short 2004). "Say" was confirmed to be the most frequent and general verb of saying for quoting and reporting in fictional dialogues in English. Its counterpart in Brazilian Portuguese, "dizer" ("say"), was also confirmed as the most frequent verb, thus corroborating most of the studies on speech presentation in translation involving this language (Pagano 2002; Cruz 2003; Alves 2006; Jesus and Pagano 2007).

The types of shifts identified support findings in the literature for different language combinations (Tannen, 1986; Klaudy 2003, Puurtinen 2003; Klaudy and

Károly 2005; Pipalová 2012). These shifts point to lower frequency in target texts for the general verb “say”, this being omitted altogether or rendered by means of verbs specifying speech function or adding a circumstantial feature or event specifying manner of saying. Lexical variety was also found to be greater in some of the target texts, particularly first translations, while retranslations tended to show a pattern closer to source texts.

Increased lexical variety of verbs of saying in first translations can be accounted for in terms of a reduction of general verbs such as “say”, but also as an attempt on the part of the translators to explore the potential of reporting verbs as what Gregory “invitations to an auditory experience” (Gregory 1967, 193).

The Conrad source text (Con_ST) presented a very different pattern in the choice of verbs of saying from Hammett and McCoy, with a wide variety of different lexical verbs and a low number of occurrences of “say”. This was found to bear an impact on both target texts. The particular profile of the Conrad source text was posited to be related to its affiliation to a literary genre in particular, different from Hammett and McCoy.

Hammet’s first translation (Ham_TT1) also showed a very distinct profile regarding translation equivalence patterns, which was accounted for in terms of there possibly being a more idiosyncratic profile for the translator, Monteiro Lobato, earlier pointed out by Alves (2006).

With regard to the omission of reporting clauses, this was found to be the case when the reporting verb was the general verb “say” and their omission did not involved a higher level of difficulty for the reader to keep track of who said what in a dialogue.

In accounting for non-equivalence and shifts, a closer examination of reporting verbs in text unfolding proved fruitful, thus supporting Toolan’s (2009) proposal of approaching reporting verbs through their role in narrative progression. Potential problems arising due to cohesive relations were found to be likely explanations for some of the shifts observed.

A final remark on the impact of the study herein reported is presented in the following and last section of this chapter.

4.12 Concluding Note

This chapter opened with a remark on the pending need in translation studies to finds ways of operationalizing a theory-informed notion of context in order to integrate text and sociocultural analysis of translations. The study herein reported was meant to address this need.

A comprehensive theory of language such as systemic functional theory offered a solid framework to deal with translation equivalence and computational resources such as those offered by R proved fruitful and enlightening for the present analysis, no doubt outperforming tools available in concordancing software amply used in corpus based studies.

Shared labour between the computer and the human analyst is here proposed as a way to meet criticism of corpus-based studies for its lack of in-depth contextual analysis. The compromise reached in our study between manual annotation of samples for critical analytical categories and automatic processing of data is thus put forward as a methodological contribution to empirical studies of translation.

Acknowledgments Research funded by the National Council for Scientific and Technological Development (CNPQ) under grants No. 308652/2010-0 and 305129/2013-9 and the State Funding Agency of Minas Gerais (FAPEMIG) under grants PPM-00087-12 and PPM-00289-14. The author would like to thank researchers Andre Luiz Rosa Teixeira and Thais Guimarães, who collaborated in the annotation and querying of the corpus samples, and Arthur de Melo Sá and Rodrigo Araújo e Castro, who developed and tested the scripts used in R.

References

Alves, Daniel. 2006. Aspectos da representação do discurso em textos traduzidos: os verbos de elocução neutros. Master's thesis, Universidade Federal de Minas Gerais.

Andrade, O.D.S. 1978. O livro brasileiro: desde 1920. 2nd ed. Rio de Janeiro: Cátedra.

Axelsson, Karin. 2011. Tag Questions in Fiction Dialogue. Ph.D. diss., University of Gothenburg, Göteborg. http://hdl.handle.net/2077/24047.

Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. *Target* 12: 241–266.

Berman, Antoine. 1990. La Retraduction Come Espace de la Traduction. *Palimpsestes* 4: 1–7.

Bönnemark, Margit. 1997. The Mimetic Mystery. A linguistic study of the genre of British and American detective fiction including a comparison with suspense fiction. PhD diss. Stockholm: University of Stockholm.

Conrad, Joseph. (1939) 1971. *Lorde Jim*. Translated by Mário Quintana. São Paulo: Abril.

Conrad, Joseph. 2002. *Lord Jim: um romance*. Trans. Julieta Cupertino. Rio de Janeiro: Revan.

Conrad, Joseph. (1900) 2005. *Lord Jim*. San Diego: ICON.

Cruz, Osilene. 2003. "Harry Potter and the chamber of secrets e sua tradução para o português do Brasil: uma análise dos verbos de elocução com base na lingüística sistêmica e nos estudos de corpora." Master's thesis, Universidade Federal de Minas Gerais.

De Haan, Pieter. 1996. More on the language of fiction. *ICAME Journal* 20:23–40.

Elson, David K. 2012. Modeling narrative discourse. Ph.D., Columbia University.

Emmott, C. 1994. Frames of reference: contextual monitoring and narrative discourse. In *Advances in written text analysis*, 157–166. Routledge.

Emmott, C., A.J. Sanford, and E.J. Dawydiak. 2007. Stylistics meets cognitive science: Studying style in fiction and readers' attention from an interdisciplinary perspective. *Style* 41(2): 204–24.

Gambier, Yves. 1994. La Retraduction, Retour et Detour. *Meta* 39: 413–417.

Glover, David. 2003. The thriller. In *The Cambridge Companion To Crime Fiction*, 135–153. ed. by Martin Priestman.

Gregg, Noel, Susan A. Sigalas, Cheri Hoy, Joseph Wisenbaker, and Carolyn McKinley. 1996. Sense of audience and the adult writer: A study across competence levels. *Reading and Writing: An Interdisciplinary Journal* 8(1): 121–137.

Gregory, Michael. 1967. Aspects of varieties differentiation. *Journal of Linguistics* 3(2): 177–198.

Hallewell, Lawrence. 1982. *Books in Brazil: a history of the publishing trade*. London: The Scarecrow Press.

Halliday, M.A.K. 1992a. Language as system and language as instance: The corpus as a theoretical construct. In *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, ed. Jan Svartvik. Berlin: Mouton de Gruyter.

Halliday, M. A. K. 1992b. Language Theory and Translation Practice. *Rivista internazionale di tecnica delta traduzione* 15–25.

Halliday, M. A. K. 2005a. Computational and quantitative studies. London: Continuum. (The collected works of M. A. K. Halliday, v. 7).

Halliday, M.A.K. 2005b. Linguistics and machine translation. In *Computational and quantitative studies*, ed. M.A.K. Halliday. London: Continuum.

Halliday, M.A.K., and C.M.I.M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*, 4th ed. Oxford: Routledge.

Hammet, Dashiell. 2002. *O homem magro*. Trans. Rubens Figuereido. São Paulo: Companhia das Letras.

Hammet, Dashiell. (1936) 1984. *A ceia dos acusados*. Trans. Monteiro Lobato. São Paulo: Abril.

Hammet, Dashiell. (1935) 1992. *The thin man*. New York: Vintage Books.

Haussen, Doris Fagundes. 1992. Rádio e política: tempos de Vargas e Perón. PhD diss., Universidade de São Paulo.

Jesus, Silvana M. D. 2004. Representação do discurso e tradução: padrões de textualização em corpora paralelo e comparável. Master's thesis, Universidade Federal de Minas Gerais.

Jesus, Silvana M. D. 2008. Relações de tradução: SAY/DIZER em corpora de textos ficcionais. PhD diss., Universidade Federal de Minas Gerais.

Jesus, S. M., Pagano, A. S. 2007. Probabilistic grammar in translation Proceedings of the 33rd International Systemic Functional Congress. São Paulo: LAEL—PUC-SP, 428–448.

Ji, Meng, ed. 2014. Empirical translation studies; Interdisciplinary methodologies explored. Sheffield; Bristol: Equinox.

Ke, Shih-Wen. 2012. Clustering a translational corpus. In *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research,* ed. Michael P Oakes and Meng Ji. Amsterdam: John Benjamins.

Klaudy, K. 2003. Languages in translation. Lectures on the theory, teaching and practice of translation. With Illustrations in English, French, German, Russian and Hungarian. Budapest: Scholastica.

Klaudy, Kinga and Krisztina Károly. 2005. Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures* 6(1).

Lajolo, M., and R. Zilberman. 1996. *A formação da leitura no Brasil*. São Paulo: Ática.

Leech, Geoffrey N., and Michael H. Short. 1981. *Style in fiction: A linguistic introduction to literary fiction*. London: Longman.

Lobato, José B.M. 1944. *A Barca de Gleyre: 40 anos de correspondência literária entre Monteiro Lobato e Godofredo Rangel*. São Paulo: Companhia Editora Nacional.

Martin, G. 1989. *Journeys through the labyrinth; Latin American fiction in the twentieth century*. London, New York: Verso.

Martins, W. 1979. História da inteligência brasileira. São Paulo: Cultrix/EDUSP. v.7.

Martínez Vilinsky, Bárbara M. 2014. Analysis of the different features and functions of dialogue in a comparable corpus of crime novels. In *The voices of suspense and their translation in thrillers*, ed. Susanne M Cadera and Anita Pavić Pintarić. Amsterdam-New York: Rodopi.

Matthiessen, Christian M. I. M. 2001. The environments of translation. In *Exploring translation and multilingual text production, beyond content*, 41–124, ed. Erich Steiner and Colin Yallop. Berlin/New York: Mouton de Gruyer.

Mccoy, Horace. 2000. *A noite dos desesperados*. Sá Editora: Translated by Renato Pompeu. São Paulo.

Mccoy, Horace. (1935) 1995. They shoot horses, don't they? London; New York: Serpent's Tail.

Mccoy, Horace. (1947) 1982. Mas não se mata cavalo? Trans. Érico Veríssimo. São Paulo: Abril Cultural.

Miceli, S. 1979. *Intelectuais e classe dirigente no Brasil (1920-1945)*. São Paulo: DIFEL.

Milton, John. 2002. *O Clube do Livro e a Tradução*. Bauru, SP: EDUSC.

Munday, Jeremy. 2002. Systems in translation: A systematic model for descriptive translation studies. *Crosscultural transgressions. Research models in translation II: Historical and ideological issues* 76–92, ed Theo Hermans. Manchester: St. Jerome.

Oakes, Michael P., and Meng Ji (eds.). 2012. *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research*. Amsterdam: John Benjamins.

Pagano, A. S. 2001. Something called books: translations and publishers′ collections in the editorial booms in Argentina and Brazil from 1930 to 1950. *Crop (FFLCH/USP)* 6:171–194.

Pagano, A.S. 2002. Translation history in Latin America: a corpora-based analysis of the strategies used by a literary translator during the 1930-1950s Brazilian publishing boom. *Cadernos de Tradução (UFSC)* 9:131–151.

Pagano, A. S., Lukin, A., Figueredo, G. 2014. Modelling proximity in a corpus of literary retranslations: a methodological proposal for clustering texts based on systemic-functional annotation of lexicogrammatical features. In *Empirical translation studies; Interdisciplinary methodologies explored*. ed Meng Ji. Sheffield; Bristol: Equinox.

Papoloski, Koskinen. 2010. Reprocessing texts: the fine line between retranslating and revising. Across Languages and Cultures, 11(1):29–49.

Pipalovâ, Renata. 2012. Framing direct speech: Reporting clauses in a contrastive study. *Prague Journal of English Studies* 74–107.

Puurtinen, Tiina. 2003. Genre-specific features of translationese? Linguistic differences between translated and non-translated finnish children's literature. *Literary e Linguistic Computing* 18(4): 389–406.

R Core Team. 2014. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Rimmon-Kenan, S. 2005. *Narrative fiction: Contemporary poetics*, 2nd ed. London: Taylor & Francis.

Rivera, J. B. 1980/86. El auge de la industria cultural (1930–1955). In *Historia de la literatura argentina*, 577–600. CEAL.

Semino, E and M Short. 2004. Corpus stylistics: A corpus-based study of speech, thought and writing presentation in narratives. Routledge.

Stape, J. H. 1996. Lord Jim. In *The Cambridge companion to Joseph Conrad*, 63–80. Cambridge: Cambridge University Press.

Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. In *Direct and indirect speech*, 311–360, ed. Florian Coulmas. DE GRUYTER MOUTON.

Tompkins, Virginia, Ying Guo, and Laura Justice. 2013. Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing* 26: 403–429.

Toolan, M. 2001. *Narrative: A critical linguistic introduction*, 2nd ed. New York: Routledge.

Toolan, Michael. 2009. *Narrative progression in the short story: A corpus stylistic approach*. Amsterdam: John Benjamins Publishing Company.

Vanderschelden, Isabelle. 2000. Why retranslate the French classics? The impact of retranslation on quality. In *On translating French literature and film II*, ed. Myriam Salama-Carr. Amsterdam: Atlanta.

Venuti, Lawrence. 2003. Retranslations: The creation of value. *Bucknell Review* 47.

Veríssimo, E. 1973. *Um certo Henrique Bertaso: pequeno retrato em que o pintor também aparece*. Porto Alegre: Globo.

Weber, R. 2008. The Shape of Direct Quotation. *The Reading Teacher* 61: 558–564.

Winters, Marion. 2007. F. Scott Fitzgerald's Die Schönen und Verdammten: A Corpus-based Study of Speech-act Report Verbs as a Feature of Translators' Style. *Meta: journal des traducteurs/ Meta: Translators' Journal* 52(3): 412–425.

Wirf Naro, Maria. 2012. Fictive orality and formality as a translation problem. In *The Translation of Fictive Dialogue*, 100–118. (Approaches to Translation Studies 35), edited by Jenny Brumme and Anna Espunya. Amsterdam; New York: Rodopi.

Chapter 5
A Corpus Analysis of Translation of Environmental News on BBC China

5.1 Introduction

BBC China represents a complex news sourcing and reporting system. Sponsored by the BBC with a special focus on Chinese-speaking communities worldwide, it includes news materials collected from a wide variety of sources in original Chinese or in Chinese translations which are mainly rendered from English. An important source of news published on BBC China are selected translations of original English news published on the website of BBC UK. The appeal of BBC China to Chinese-speaking audiences, especially those residing inside of mainland China, consists in its perceived impartial and balanced view over domestic and international issues, especially those related to China. BBC China is one of the few major western news outlets that have a wide presence in the country, as most western reporting services are subject to strict media regulation and control from the central government.

The British news outlet provides an important alternative news source for those with limited access to original English news, particularly readers who can only read in the Chinese language, despite the increased level of English literacy among younger generations. More importantly, it serves as a powerful tool to promote western ideas and values as well as inform a large and rapidly growing Chinese-speaking audience globally. Regional versions of BBC such as BBC China may be seen as a leader in the provision of localised news materials to target audiences amidst the latest waves of globalisation which give leverage to emerging digital technologies and media resources to facilitate cross-cultural and cross-language communication.

There is a growing body of research focusing on environmental media, especially cross-national studies which intend to establish the nexus between variations in news reporting and explanatory factors that lie beyond national borders. The current study builds upon this growing trend of research by focusing on a particular type of trans-national environmental reporting, i.e. translation and adaption of

S. Laviosa et al., *Textual and Contextual Analysis in Empirical Translation Studies*, New Frontiers in Translation Studies,
DOI 10.1007/978-981-10-1969-2_5

environmental news for non-English speaking audiences. This is exemplified by BBC China whose reporting system exhibits a hybridity of news sources including those originally published in Chinese as well as translations and adaptations of English news. As a regional and largely localised website of BBC UK, the popularity of BBC China with a well-targeted audience underscores, on the one hand, the rise of multilingualism and, on the other hand, the influence and values of original English reporting which shine through the production of news materials published on BBC China.

The popularity of BBC China and its impact on global Chinese-speaking audiences cannot be explained in a uni-dimensional manner, i.e. its intrinsic connection with BBC UK. This is because, despite the endorsed reporting strategies of original English news, a literal translation of original English news would necessarily receive resistance and criticism from the global Chinese-speaking readership with an extremely complex ethic, cultural, ideological and age structure. What are the reporting and editorial strategies developed by BBC China to inform its language-specific readership, especially about relatively new problem areas of global concern like climate change and environmental protection? Does BBC China global reporting of such issues fit in with prevalent Western reporting frameworks discussed in previous studies (Boykoff 2007; Carvalho 2007; Dirikx and Gelders 2010), especially with regard to the translation and adaption of original English news?

The current study explores these intriguing issues and aims to develop new analytical instruments to provide a useful focal point for global environmental reporting. The methods of analysis draw upon those widely used in corpus-based translation studies, environmental media, contrastive linguistics and comparative area studies. The highlight of translation and adaptation of news underpins the growing importance attached to localised digital media amidst the latest waves of globalisation and cross-lingual communication (Franklin 2014; Rooke 2013). Translation represents a particular type of discourse and language variety, which differs from original writing. By definition it involves two languages and socio-cultural systems, known as the source language and the target language (Toury 2012). The complexity of translation lies in the development of communication strategies to facilitate the understanding and interaction between two distinct language and cultural systems.

With the rise of descriptive translation studies in the 1990s (Toury's polysystem theory and Bassnett's translation turn in cultural studies) (Toury 1995; Bassnett 1998), which is essentially social and culturally oriented; and the introduction of empirical research methodologies (Baker 1995; Oakes and Ji 2012), more efforts in contemporary translation studies have been directed at exploring useful linguistic and textual patterns in translated texts. The exploration of underlying textual patterns aims to identify the features that characterize a particular type of translation genres or discourses, for example, literary translation, specialised translation of economics, finance, business, news and broadcast materials, etc. These textual features may be investigated in terms of translation stylistics (Bosseaux 2004; Malmkjær 2004; Bernardini 2005); cognitive functions of translation (Deignan and

Potter 2004; Shuttleworth 2011; Stefanowitsch and Gries 2007; Schäffner 2004); translation universals and laws (Mauranen and Kujamäki 2004; Olohan and Baker 2000; Pastor et al. 2008); contrastive linguistic features (Granger 2003; Goethals 2007; Kenny 2014); genre-specific translation features (Trosborg 1997; Scarpa 1999; Puurtinen 2003; Ji 2013a). Following this empirical line of translation research, the current study aims to establish the distinct dimensions of reporting and associated communicative strategies of mediated environmental news published on BBC China, which represents an important type of specialised translation.

5.2 Research Question

This study sets out to investigate a particular type of environmental reporting, i.e. the translation and adaptation of original English environmental news published on BBC China, which represents a hybrid repository of global news collected and re-designed for Chinese-speaking communities worldwide. Whilst there is a considerable amount of environmental news on BBC China that are collected from original Chinese sources, the focus of the current study is translated and often abridged English news from the parent website of BBC UK. The significance of translated news is that it provides global Chinese audiences with first-hand information on latest scientific discoveries and research outcomes from environmental science; debates on national and international policies regarding environmental issues, as environmental protection has become an integral part of international relations; and more importantly, critical and timely reports on the impact and consequences of climate change on national economies and communities that are most vulnerable to various types of environmental deterioration.

Environmental news assume a critical role in informing and raising awareness among a highly diversified global Chinese-speaking audience about the importance and urgency of environmental protection. This is achieved by carefully selecting, translating and adapting news from original English sources, for example, BBC UK, targeting news that are likely to be of interest to the Chinese audiences and provoke discussions and debates on the social events reported. Research has revealed that environmental news has developed different dimensions of reporting in different national contexts which are related to societal variables operating at different levels including ideology (Dirikx and Gelders 2010), politics (Carvalho 2005; Carvalho and Burgess 2005), economic and national media market privatisation (Tong 2014), and journalism (Boykoff and Boykoff 2004).

Frame analysis provides an effective research methodology to identify and configure the societal conditions and factors which shape understanding of issues such as climate change and the social significance of environmental protection. Hope (2010) outlines three levels of frames in a typical frame analysis which include primary frames, meta-frames and issue frames. Primary frames encompass natural and social frames. They provide the most general and basic level of understanding for the conceptualisation of a highlighted or chosen issue or ‘object’ as natural or a social event. Meta-frames, on the other hand, may be construed as a set of ideological, political and cultural ideas and concepts which prevail in special societal contexts during specific periods in time. Lastly, issue frames refer to linguistic and reporting strategies developed to process real-life news events.

The construction of these frames at varying levels holds the key to achieving the intended reporting effects. The robustness and thoroughness of the frames developed may be measured and assessed by their relative position within the larger social and belief system, i.e. marginalised or central; and their relevance to the realities and everyday life experiences of the targeted audience (Snow and Benford 1988). With regard to the current study, this means first, to consider how the meta-frames or underlying reporting strategies used by BBC China fit in with the social conditions and journalistic norms prevailing in the target social system. Second, it means to consider how reporting strategies developed by BBC China reflect and accommodate the expectations of its targeted Chinese-speaking audiences regarding issues revolving environment, development and their own well-being. This study builds on studies which deploy the frame analysis theory to identify consistent linguistic and textual patterns in translated environmental news published on BBC China. It will attempt to classify patterns or frames detected by the computer-assisted sentiment analysis in line with the primary frames and meta-frames described above. It will assess the validity and robustness of these frames of environmental reporting within the larger social context of mainland China. The analysis of these frames and associated reporting strategies will yield insights into the problematisation of environmental issue within the current climate of China's accelerated social reform and economic development.

5.3 Computer-Assisted Configuration of Frames in Media Analysis: An Overview of Analytical Techniques

An increasing number of frame-oriented media studies focus on the computer-assisted extraction and modelling of underlying patterns in media texts, printed or online, in an effort to develop robust and replicable measure instruments and/or scales for empirical analysis. Andersen (2007) offers a quantitative content analysis of a large-scale corpus of newspaper articles on WTO Doha Development Agenda negotiations. The goal of the corpus analysis is to uncover institutionalised meta-frames that depict multilateral trade negotiations between 2001 and 2006. In Anderson's study, meta-frames are characterized by persistent regularities in news coverage and are defined in terms of content institutionalization. Examples are the recurrent description of trade talks as furthering development and the unequal distribution of resources which underscore the lack of substantial advances in the trade talks reported during the six-year period under study. The development of meta-frames, i.e. recurrent themes and general trends in the environmental news database has largely drawn upon quantitative analyses.

Anderson uses two sets of important techniques to detect and extract meta-frames in his quantitative analysis, i.e. open and iterative thematic coding and classification of countries based on their group membership which are G20 and OECD (Organisation for Economic Cooperation and Development). These represent the interests of developing and developed countries respectively. An important feature of the open and iterative coding of news data is the conflation of description items in a successive manner to extract the main components or features that maximally explain the differences and similarities among the five countries investigated, i.e. USA, Spain, UK, India and Argentina. For example, in an effort to explore any underlying patterns of the purposes of trade reporting over the six-year period, the researcher analyses a third of the entire database manually and attributes the news articles to as many as 384 purpose descriptions. This high-dimension corpus annotation result does not achieve a clear pattern recognition. Anderson goes on to repeat the coding by using more general descriptive terms to reduce the initial dimensions of coding.

The result is a list of six different definitions which capture the main purposes of trade reporting in six countries across the six-year time span: development for the poor; development for the rich; boost and save world economy; liberalisation; international security and cooperation; and help environment. By dividing the six countries into two groups according to their economic status, i.e. G20 and OECD, clear and consistent patterns begin to emerge: the two most important purposes of trade reporting by both G20 and OECD are liberalisation and development for the poor. Whereas liberalisation enjoys a strong presence across the entire time span, the development theme has acquired more importance over time. Both purposes have thus been defined as institutionalised topics which give rise to the meta-frames of liberalisation and development. In a similar fashion, Anderson develops two sets of framing scales bearing on jeopardy and responsibility attribution.

Anderson's study despite being quantitative in nature, involves a considerable amount of manual annotation which is subject to personal interpretation. Moreover, the iterative process used to detect underlying patterns or the so-called meta-frames, which is achieved by reducing initial descriptive categories, inevitably hinders an in-depth and fine-grained analysis of the data. A fuller description of the annotation scheme used is required to explore the relationship between the features of reporting and external explanatory factors in further statistical analysis. The significance of Anderson's study is that it makes a useful initial attempt at developing transnational meta-frames based on extensive observation of large amounts of raw reporting data. This is particularly useful for developing new hypotheses and theories in emerging research areas such as media frame analysis.

An increasing number of empirical studies have focused on advancing analytical methods and techniques in the process of frame extraction. David et al. (2011) compare the results of manual annotation and frame detection, similar to Anderson's study, with the computer-assisted frame configuration based on the statistical analysis of quantitative published news data. In the process of computer-assisted frame extraction, they use a community or frame detection technique which is driven by the modularity in the data set. This method is similar to traditional clustering methods such as k-means cluster and hierarchical cluster analysis. One of the advantages of this method, as highlighted by the authors, is that it retains textual features that are not recurrent or consistent in the news data collected. Frequency lies at the heart of statistics. Low frequency data with sparse distribution in the data tend to be omitted as they are not statistically significant. David et al. argue that this practice could potentially overlook textual features that are important in the qualitative analysis, thus giving rise to criticism from researchers upholding hermeneutic and qualitative approaches to media and discourse analysis.

David et al. found that the result obtained from the computational analysis, i.e. communities or frames which classify news articles on population and health management published in the Philippines between 1987 and 2007, largely agreed with or were comparable to the result obtained from the manual or qualitative analysis. Hence, the advantage of the computer-assisted analysis over completely qualitative interpretive approaches is evident, given the cost and labour involved in the latter. The corpus analysis of mediated environmental news presented in this chapter evokes a different research philosophy from the one endorsed by David et al. Its priority is not attuning the result of statistical analysis to traditional content analysis that is done largely manually; on the contrary, the current study is essentially corpus-driven and entrusts statistical analysis to take the lead in the exploration of new patterns in the data base. The focus is on the extraction of a small set of textual features or attitudinal languages from an extensive list of automatic annotation categories. The textual features thus retrieved will be discussed in line with frame analysis, as they provide important information regarding the reporting strategies developed by BBC China.

5.4 A Frame Analysis of Translation of Environmental Reporting on BBC China

5.4.1 Collection of Translated and Adapted News on Environmental Issues in BBC China

The first stage in the study of the translation of environmental news on BBC China is the collection of relevant data from the news website. As stated from the outset, BBC China represents a hybrid news reporting system which incorporates news collected from a variety of sources, including translated and adapted news originally published on the parent website BBC UK; news collected from other national news outlets such as Nikkei News (Japan); The Australian (Australia); CNN (USA); Le Monde (France), The Moscow Times (Russia), major Chinese news outlets including China Daily, People's Daily. It is relatively convenient to detect news which are selected, translated and adapted for publication on BBC China, as the editor and the Chinese translator are often identified at the end of each news entry.

This editorial practice stands in contrast with other regional websites of BBC, for example, BBC Mundo (in Spanish), where translated news from other languages are not differentiated from news reporting by regional correspondents writing in local languages, e.g. Latin American Spanish.

The visibility given to news translated from other sources, mainly Western news outlets, as well as neighbouring countries such as Japan, Korea and Russia, indicates particular editorial strategies developed by BBC China to offer an impartial, balanced and cosmopolitan perspective of global news events to its highly diverse Chinese-speaking audiences. Also, the status of translated news as a news genre parallel to original Chinese reporting implies the value and credibility attributed to foreign news by the Chinese audiences. The automatic retrieval of translated news on environmental issues and climate change benefited from the parallel reporting system of BBC China. By inputting key research words such as global warming, climate change and environmental deterioration with a space followed by the Chinese character YI (which means translation), a number of news entries came up. The automatically returned news entries were then screened manually based on the levels of relevance to this study. News entries which discussed marginally the input key words were excluded to streamline the analysis.

More weight was given to news entries which dealt specifically with the social impact and implications of environmental issues. Discussions on unobtrusive issue-areas such as climate change and global warming were deliberately included in the news search alongside investigative reporting on environmental deterioration, which prevail in Chinese environmental media studies. The inclusion of news reporting on climate change aims to explore how foreign news outlets such as BBC China foster an increased social awareness and scientific attitude towards environmental phenomena such as global warming among Chinese audiences, given that similar issues tend to receive much less attention in national news reporting in China. The result of the automatic retrieval followed by the manual filtering gave a relatively small yet representative list of environmental news articles published on BBC China in the last five years, i.e. from 2009 to 2014.

The total number of news articles, exclusively translated and abridged news and directly related to environmental issues, was eighty-seven with the running tokens of each article ranging between 400 and 2000 character words. The manually filtered environmental news were then transcribed onto text files and labelled individually. Based on the thematic topics covered, these raw news materials were then classified into three recurrent categories which are (1) governance; (2) international relations and (3) environmental science. The first type of environmental news, i.e. governance deals with the social impact of environmental issues at a national level. It is not limited to the management and handling of environmental problems by the Chinese government and its local representatives but extends to include reporting on national policies, legislation and high-profile events of environmental protection in other parts of the world, especially in the USA, UK, Russia, India and Japan. The distinctly wide breadth of reporting on environmental governance at national level enables audiences to develop their critical understanding and evaluation of the effectiveness and deficiencies of national environmental policies within the global context.

The second type of environmental news, i.e. environment and international relations underlies the international dimension of environmental media reporting. In the transcription and classification of translated news published on BBC China, it became clear that environmental issues were often framed within a highly dynamic international setting. Similar to Anderson's study of the reporting of WTO trade negotiations, environmental issues provide an important object of debate and division between developing and developed countries in terms of responsibility attribution and distribution of resources to combat environmental deterioration. This second type of environmental news represents an integral part of environmental media reporting by BBC China. It informs and prepares a globally minded Chinese-speaking audience, especially among younger generations. Which look at environmental issues beyond national borders.

The last type of environmental news that recurred in the translation data collected from BBC China may be put under the general category of environmental science. The environmental news included in this category dealt,, amongst other topics, with the latest scientific breakthroughs; the advances made in the use of new energies on an industrial scale; the impact of extreme weather conditions on national economies and people's daily lives. If the first two types of environmental news are seen as primary frames in the social domain, the third type of environmental news is essentially a natural frame which engages the public in the scientific and scholarly debates of climate change and global warming. This type of environmental news receives less attention in national news reporting inside of China, as the majority of the general public accepts, to varying degrees, climate change and global warming as natural results of environmental deterioration which come at the cost of the country's economic development.

The everyday experience of severe air and water pollution makes few people doubtful of the anthropogenic nature of environmental changes. The psychological predisposition of the public in China stands in stark contrast with some western countries, for instance, USA and Australia, where the (quasi-) scientific nature of climate change and global warming as depicted by national mainstream media outlets is still very much in the balance. However, the situation in China represents a dilemma: whereas most people agree on the widespread existence of industrial pollution and climate change, few people take real action to deal with the situation. It would be interesting to investigate, through the analysis of translated environmental news data, how BBC China frames news of environmental science selected from original western media sources to inform its Chinese-speaking audiences on the severe impact of environmental changes on the natural world; the availability of effective scientific tools and methods to improve environmental conditions; and perhaps, more importantly, to mobilise the public to take actions to prevent further environmental deterioration.

5.4.2 *Development of Analytical Instruments for Mediated Environmental Reporting on BBC China*

5.4.2.1 Sentiment Analysis

The sentiment analysis tool used in the current study was developed by the Centre for Computer Corpus Research on Language of the University of Lancaster, UK. It is known as USAS. The development of USAS was originally based on McArthur's *Longman Lexicon of Contemporary English* (McArthur 1992). USAS has a multi-tier structure which covers twenty-one major discourse fields and domains which are labeled alphabetically. Within each field, subdivisions are provided based on the sentiment (especially subjective and evaluative) properties of terms and expressions classified in each major domain. The Chinese version of USAS (CH-USAS hereafter) draws upon the English taxonomy of sentiment analysis described above. It is used in the annotation of translated environmental news published on BBC China.

The development of CH-USAS was based on the Linguistic Data Consortium English-Chinese translation lexicon (Huang and Graff 2002). The automatic annotation results have some important features which illustrate the developing nature of CH-USAS. Similar to USAS, sentiment properties attributed to a Chinese character word (for the detection of legitimate character words from natural language strings, CH-USAS uses the Chinese segmentation tool known as Stanford Chinese Word Segmenter) often cover several discourse fields. Here is a typical example of automatically annotated words based on the sentiment analysis of its use in context:

预 测 (**Yu-Ce**) (**predict**)

Automatic sentiment analysis result:

X9.1-/S2mf S1.2.6-/S2mf N2 X2.4 Q2.2/T1.1.3 X2.6+/T1.1.3/S9 X6/A7− N2/A7−

Detailed explanation of each sentiment tagging subdivision:

X9.1−: Terms depicting (level of) ability/intelligence such as DE-SKILLED (−), DINNY (−), NOT MUCH OF A (−), BLIND LEADING THE BLIND (−).

S2mf: Term indicating that particular words relate to/denote people, m: male; f: female

S1.2.6−: Terms depicting (level of) sensibleness/absurdity such as ABSURD (−), BALONEY (−), CRAZILY (−), IDIOTIC (−), LUDICROUS (−), NAÏVE (−), NONSENSICAL (−)

N2: Mathematical terms

X2.4: Terms relating to investigation/examination such as ANALYTIC, ANALYSE, ASSESSED, CONSIDERED, CONTEMPLATE, DELVING

Q2.2: Speech acts terms such as ARGUE BACK, ASK AROUND, ACCOUNT, ADDRESSING, ADVOCATE

T1.1.3: General terms relating to a future (period/point in) time such as DEFER, FUTURE, HEREAFTER, IMMINENT, IMPENDING, SOMEDAY, TOMORROW, WILL, COMING

X2.6+: Terms depicting (level of) expectation such as ANTICIPATED (+), ENVISAGE (+), EXPECTANCY (+), FORECAST (+), FORESEE (+)

S9: Terms relating to religions and the supernatural

X6: Terms relating to decisions/decision making

A7−: Abstract terms of modality (possibility, necessity, certainty, etc.)

As the detailed breakdown of the automatic sentiment analysis shows, CH-USAS has attributed eleven sentiment categories to the character word Yu Ce, which means prediction/predict. Among the eleven sentiment tagging categories, six are directly related to the character word: S2mf; X2.4; T1.1.3; X2.6+; X6 and A7−. The other five categories (S9; Q2.2; N2; S1.2.6−; X9.1−) are marginally related or not relevant in this specific context. This has to do the bilingual lexicon (LDC translation lexicon) used in the development of the CH-USAS system, as a Chinese character word may have various English translations in different contexts. To increase the recall and precision rate, after the automatic tagging performed by CH-USAS, two human annotators went through the annotation results and manually filtered out tagging categories which were not relevant in the news context. This was a time-consuming process, as the qualification process required comparing the tagging categories against the textual context of the news to retain the most relevant items. As the pruning process inevitably involves an element of subjectivity, the two human annotators compared their rating results (relevant tagging items were marked positive and non-relevant items were marked negative) at the end of the manual annotation. The kappa statistics used to assess inter-rater agreement was 0.84 which was high enough to warrant the validity of the post-editing results.

The advantages of using automatic sentiment analysis tools like USAS and its various language versions (USAS is operational in various local versions, for example, Chinese, Italian, Dutch, Portuguese and Spanish)[1] are multiple. Firstly, it represents a useful first attempt at experimenting with a computer-assisted annotation system, which, despite its developing nature, is valid with very different languages, i.e. from languages using the Latin alphabet to typical Asian languages. This helps to overcome technical and linguistic difficulties involved in cross-lingual and cross-cultural analysis. USAS takes advantage of the cognitive mechanisms that underlie many modern languages and represents a higher level language processing system, since other computer analytical systems such as part-of-speech tagging and syntactical parsing tend to be language-dependent, especially with non-cognate and/or typologically distinct languages. The deployment of an annotation tool that is valid across languages and genres makes an important contribution to the development of theoretical and hypothetical models with a wide

[1]For more details, see http://ucrel.lancs.ac.uk/usas/ (last access on 25 October 2014).

application in different regional and national contexts so as to fill in a critical gap in comparative environmental media analysis.

Secondly, the use of automatic corpus annotation tools like USAS is instrumental in processing larges-scale data bases to extract important quantitative information. As Anderson's study and an increasing number of quantitative media studies show, the statistical processing and modelling of media data is a pre-requisite to conceptualising and developing theoretical constructs such as meta-frames and measure scales for empirical media studies. With a small set of news events as the base of investigation, the result obtained or the conclusion drawn are necessarily circumscribed and insufficient, as required by many statistical analyses, to yield insights into the relation such as co-variation, changing dependence relationship between social and cultural factors, moderating and mediating variables, and the framing and editorial strategies devised for environmental media reporting purposes.

After collecting the translations of environmental news from the website of BBC China, the raw news data were first inputted into the automatic tagging system of CH-USAS. Next, they underwent a thorough manual screening to retain tagging items which were deemed relevant in the news context. Given the intensive efforts required in the checking process, we focused on the compilation of a relatively small corpus of translated environmental news for the development of a set of exploratory meta-frames on the reporting strategies of BBC China over a five-year period from 2009 to 2014.

5.4.2.2 Research Aims and Design

Before proceeding to the statistical analysis, it is important to elaborate on the purpose of the quantitative analysis that follows the collection of translated news data. There are two main types of statistics that are particularly relevant to the current study, i.e. pattern detection and feature discrimination for classification purposes. Similar to Anderson's study of WTO trade negotiations through the development of meta-frames, the first aim of our study is to identify reporting strategies used by BBC China. This is based on the quantitative analysis of recurrent textual patterns in translated environmental news. The recurrent textual patterns retrieved are then defined as meta-frames. These meta-frames reveal how the selection and translation of environmental news enables BBC China to inform and mobilise its targeted Chinese-speaking audiences about environmental issues and the social ramification of climate change events.

The second aim of this study is to highlight textual features which effectively discriminate different types of environmental news. This involves the identification of a small set of textual features drawing upon the result of sentiment analysis to explain differences among the three sub-genres of translated environmental news in the current study, i.e. governance, international relation and environmental science, as well as intra-genre variations. As explained in Sect. 5.4.2.1, the current version of USAS has twenty-one major discourse fields and two hundred and thirty-two category labels. On the one hand, the comprehensive annotation system helps to increase the depth and scope of linguistic and sentiment analysis. On the other hand, it poses difficulties for the detection of textual features characteristic of each of the three environmental news genres.

To achieve our first objective, exploratory factor analysis (EFA) was applied in the multidimensional annotated data. EFA is a widely used statistical technique which configures the underlying patterns of a large set of observed variables by classifying the original variables into a reduced set of variables known as principal components, factors, conceptual dimensions or measurement scales. Re-classified observed variables are either closely or loosely related to the newly constructed dimensions as indicated by their varying component loadings. A large positive loading indicate a strong correlation between the observed variable and the factor. A moderate positive loading implies the limited contribution of the variable to the hosting factor. Component loading scores may also have negative values and the interpretation is slightly different. A large negative loading suggests that an increase of value of the observed variable will cause a proportional decrease in the value of the dependent variable.

In our study, if some linguistic features were negatively related to a conceptual scale built by EFA, an increase of frequency of these linguistic features was linked up with the decrease in the frequency of occurrence of environmental news sub-genres which were highly related to that particular conceptual dimension. By eliminating observed variables with low loadings on the conceptual dimensions and low item-total correlation scores within each dimension, EFA can effectively extract textual features from a large number of original observed variables to predict the behaviour of the dependent variable. The small set of textual features retained are then used to discriminate and classify under-specified environmental news, which is the second objective of our study. Confirmatory statistics Discriminate Analysis (DA) was used to assess the validity and efficiency of the conceptual scales or analytical models built by EFA.

5.4.2.3 Reduction of High Dimensions of Textual Features for Sentiment Analysis

Environmental media research has developed rapidly in terms of the scope and depth of analysis, with an increasing number of cross-national and longitudinal case studies. However, current studies, especially those in line with framing analysis, have focused on reporting news on environmental events sharing similar thematic content. Examples are the debates and consensus building around climate change and global warming (Boykoff 2007); investigative reporting on industrial pollution and corporate responsibilities (Tong 2014); international collaboration on environmental protection and economic development (Janetos et al. 2012; Boykoff 2008). The wide range of discussions on environmental issues reflect competing priorities, distinct agenda setting and issue framing revolving around investment on environment in different national contexts. This has given rise to a variety of sub-genres of environmental reporting including, among others, efficiency and accountability of governments for pollution regulation, international collaboration on environment protection and development as well as environmental science. It is worth pointing out that very limited research has been conducted on the differences and similarities among these emerging sub-genres of environmental media reporting.

This study is the first to investigate the contrastive textual features of three sub-genres of environmental media, i.e. environmental governance, environment and international relations, and environmental science. These sub-genres represent the bulk of mediated or translated environment news published on BBC China. Exploratory statistical techniques like EFA are particularly useful in the absence of well-developed and solid theoretical frameworks to inform purposely designed empirical analyses like the one carried out in the present study. Based on the result of sentiment analysis and manual editing, an EFA was run on the remaining observed variables, i.e. CH-USAS tagging categories. Table 5.1 shows part of the iterative search and selection procedure that led to the detection of less than twenty sentiment analysis categories out of the initial two hundred tagging items of CH-USAS.

These highlighted items are deemed most relevant in the analysis of mediated environmental news given their computed Kaiser-Meyer-Olkin (KMO) scores shown in Table 5.1. KMO scores measure the appropriateness of the data gathered for factor analysis. The KMO index ranges from 0 to 1: in the 0.90 s as marvelous, in the 0.80 s as meritorious, in the 0.70 s as middling, in the 0.60 s as mediocre, in the 0.50 s as miserable, and below 0.50 as unacceptable (Tabachnick and Fidell 2007; Williams et al. 2012). As we can see in Table 5.1, the KMO scores of all of the remaining sentimental analysis categories range between 0.7 and 1.0. This indicates the proportion of variance in the variables is caused by underlying factors, thus it allows for the application of factor analysis. A large number of sentiment analysis categories have been removed from this initial screening due to missing value and/or detected lack of sampling adequacy.

Based on the result of KMO scores, EFA constructed a model which separates the remaining nineteen sentiment analysis categories into two dimensions. Table 5.2 shows the variable and factor correlation scores between the observed variables and the conceptual scales constructed. A large positive score is a sign of the variable as a good measurement of the scale, as it indicates the strong

Table 5.1 Kaiser-Meyer-Olkin measure of sampling adequacy

1	A1.5.1	0.722	8	F1	0.781	15	X2.1	0.905
2	A13.3	0.909	9	K6	0.730	16	X9.1	0.822
3	A13.6	0.794	10	L1	0.830	17	Z2.5	0.730
4	A17	0.893	11	O1.3	0.812	18	Z7	0.924
5	A4.1	0.853	12	O3	0.894	19	Z99	0.813
6	A5	0.715	13	S7.4	0.802		KMO	0.812
7	B3	0.810	14	T 1.2	0.892			

Table 5.2 Variables/factor correlation

No	Observed variable	F1	F2
1	A1.2	0.745	−0.334
2	A1.6	0.281	−0.133
3	A1.7	0.321	−0.031
4	A4.2	−0.244	0.060
5	A5.2	0.383	−0.087
6	A5.4	0.251	0.070
7	A6	0.286	−0.151
8	A15	0.165	−0.059
9	E3	−0.079	−0.188
10	E4.2	0.346	−0.054
11	E4.1	−0.104	0.324
12	S 1.2.6	0.140	0.692
13	S 4.3.2	−0.208	−0.011
14	S3.1	−0.071	0.175
15	S6	−0.053	−0.039
16	S7.3	−0.200	0.128
17	S8	−0.227	0.081
18	Z4	−0.099	−0.248
19	Z7	0.109	0.467

correlation between the variable and the conceptual scale constructed. A large negative score also provides important information regarding the distribution or variation in the dependent variable as a result of variation in the observed variable. The efficiency of the model is enhanced by removing variables with ambiguous or equally small loadings on both dimensions, i.e. A15, E3, S4.3.2, S3.1, S6, S7.3, S8, and Z4. Table 5.1 shows the start of an iterative process which deletes successively observed variables that have small variable/factor correlation scores and the details will not be repeated here.

After the iterative selection process, a greatly streamlined analytical model emerged. This encompassed only four sentiment analysis categories, i.e. A1.2 (General/abstract terms relating to appropriateness, suitability, aptness); A1.6 (General/abstract terms denoting (level of) practicality/abstraction); A5.2 (Evaluation: true/false) and E4.2 (Level of contentment). Tables 5.3 and 5.4 show the internal structure of the two dimensional model built by EFA. The first dimension is substantiated by A1.6 and E4.2, whereas the second dimension includes A1.2 and A5.2. Table 5.4 displays the eigenvalues of the two factors. Eigenvalue is an important measurement in factor analysis. As a rule of thumb, factors with eigenvalues greater than one tend to be retained in EFA, whereas those with values lower than one are normally discarded. In the current study, both factors were retained to assist with the discrimination of sub-genres of environmental news reporting at the next stage (Fig. 5.1).

Table 5.3 Correlations between variables and factors after Varimax rotation

		D1	D2
E4.2	Level of contentment	0.991	−0.172
A1.6	General/abstract terms denoting (level of) practicality/abstraction	0.547	0.164
A1.2	General/abstract terms relating to appropriateness, suitability, aptness	−0.149	0.969
A5.2	Evaluation: true/false	0.275	0.496

Table 5.4 Eigenvalues

	F1	F2
Eigenvalue	1.263	0.829
Variability (%)	31.573	20.727
Cumulative (%)	31.573	52.300

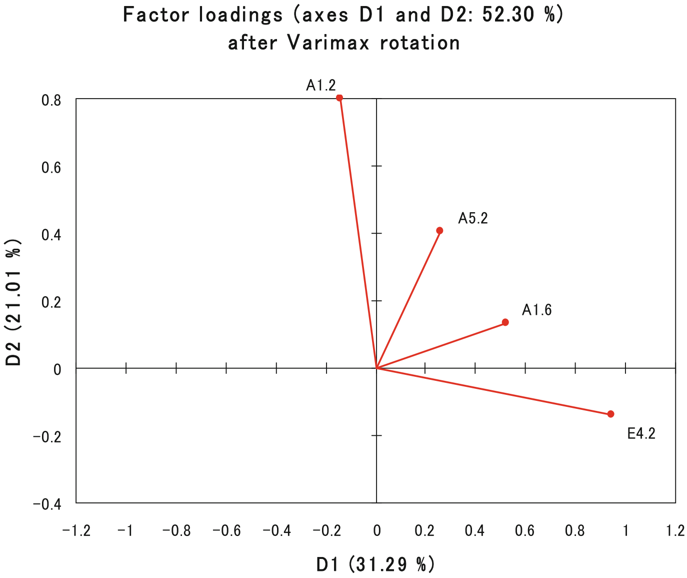


Fig. 5.1 Factor loadings

5.4.2.4 Discriminant Analysis

In an effort to test the efficiency of the observed variables in differentiating un-classified environmental news, discriminant analysis (DA) was used. In statistical analysis, the selection of relevant tests draws upon a number of factors, for example, the level of measurement, i.e. nominal, ordinal, interval, and ratio. In the current study, the dependent variable is categorical and no-metric. This implies DA should be used instead of multiple regression analysis (MRA), which is another widely used group membership prediction technique, as DA classifies cases of the dependent variable into non-interval outcomes, for example, different sub-genres of translated environmental news in the current study. DA predicts variations and changes in the dependent variable based on a linear combination of the values of interval explanatory variables.

DA assumes the normal distribution of sampled data and linear relationship among explanatory variables, the independence among predictor variables is critical to the use of DA for membership prediction and discrimination purposes. DA assumes independence of cases and multivariate linearity of relationships among the independent variables in each category of the dependent variable. Multicollinearity or high correlations among independent variables can be a problem. Tolerance and Variance Inflation Factor (VIF) are very popular as two multi-collinearity diagnostics. The threshold value of VIF to determine the existence of high correlation among the independent variables is normally set at 10 (Aline 2010). Tolerance value ranges between zero and one. The smaller the tolerance value, the higher the level of multi-collinearity. That is, tolerance close to 1.0 indicates the predictor variable in question is not redundant with other predictors already in the regression equation, while a tolerance close to zero indicates a high degree of redundancy. Table 5.5 shows the result of the multicollinearity statistics which serves as an indication of the adequacy of four remaining sentiment analysis categories for the following DA. Given that their VIF values are much smaller than 10 and their tolerance values are close to one, they are deemed appropriate to use for membership attribution purposes.

Table 5.6 shows the result of DA in which the four sentiment analysis categories were used to predict the group membership of fifty randomly selected articles from the translated environmental news data base constructed. Figures on the diagonal are cases which have been rightly predicated by the DA model (20 for subgenre 1; 8 for subgenre 2 and 7 for subgenre 3); and their percentage of total, which includes both correctly and incorrectly classified cases are presented in the last column (% Correct). As can be seen in Table 5.6, the DA model rightly predicated almost ninety per cent of the total cases of sub-genre 2 (international relations and environment); the success rate remains at a high level of 80 % with the discrimination of sub-genre 1 (environment and governance); and the lowest success rate was just above two fifths (44 %) with the identification of sub-genre 3 (environmental science). As a result, the overall accuracy of membership attribution of the DA model constructed is 71.1 %.

Table 5.5 Multicollinearity statistics

Statistic	A1.2	A1.6	E4.2	A5.2
Tolerance	0.800	0.746	0.734	0.798
VIF	1.251	1.340	1.363	1.253

Table 5.6 Confusion matrix

Membership	Sub-genre	1	2	3	Total	Correct (%)
1	Governance (China, Japan, US, Africa, Australia)	20	4	1	25	80.00
2	International relations and environment	0	8	1	9	88.9
3	Environmental science	5	4	7	16	44.5
Total		25	16	9	50	71.1

The breakdown of the accuracy rates shows the efficiency of the DA model with the classification of sub-genre 1 and sub-genre 2; and its limitation with the isolation of sub-genre 3 from the rest of the news texts. This suggests that with translated environmental news from other sources, for example, BBC UK, BBC China tends to deploy consistent translation and adaptation strategies with the reporting of issues revolving two sets of themes. These are (1) relation between (in) effective governance and environmental deterioration, for example, air pollution in China and mismanagement of nuclear plants in Japan; and (2) the problematisation and prioritisation of climate change, global warming and other environmental issues on agendas of international gatherings and organisations as well as part of national policies for and approaches to internationalisation. The low accuracy rate with the membership attribution of sub-genre 3 of environmental science indicates that the DA model we constructed does not help much with the study of translated environmental science news. The next section will examine the underlying patterns of the three types of translated environmental news, in an effort to identify essential reporting strategies developed by BBC China to frame and problematise environmental issues from various angles and perspectives and to inform its well-targeted Chinese-speaking audiences.

5.5 Research Findings

Table 5.7 shows the result of the DA, which allows a better understanding of the reclassification process. The first column shows the number of the fifty randomly selected news items from the translated environmental news data base. The column Prior shows the manual labelling of the news texts: 1 for sub-genre 1 (governance and environment); 2 for sub-genre 2 (international relations and environment) and 3 for sub-genre 3 (environmental science). The column Posterior shows the automatic reclassification of the news texts based on the DA model. A disagreement between the prior category and the posterior category indicates a misclassification. The accuracy and reliability of the manual labelling, as explained earlier, is warranted by a high inter-rater validity score. The two columns F1 and F2 display the factor

Table 5.7 Prior and posterior classification and factor scores

Observation	Prior	Posterior	F1	F2
Obs12	1	2	−1.192	−0.046
Obs34	1	2	−1.192	−0.046
Obs48	1	2	−1.192	−0.046
Obs37	1	2	−1.452	−0.082
Obs39	1	3	−0.452	0.182
Obs7	1	1	0.288	0.411
Obs13	1	1	0.288	0.411
Obs27	1	1	0.288	0.411
Obs38	1	1	0.288	0.411
Obs45	1	1	0.288	0.411
Obs50	1	1	0.288	0.411
Obs1	1	1	1.028	0.440
Obs17	1	1	1.028	0.440
Obs32	1	1	1.028	0.440
Obs44	1	1	1.028	0.440
Obs46	1	1	1.481	0.454
Obs6	1	1	1.768	0.469
Obs26	1	1	1.768	0.469
Obs33	1	1	1.768	0.569
Obs31	1	1	2.221	0.565
Obs36	1	1	2.741	0.522
Obs40	1	1	2.674	0.599
Obs29	1	1	3.247	0.527
Obs30	1	1	3.247	0.627
Obs35	1	1	3.987	0.655
Obs2	2	2	−1.192	−0.046
Obs4	2	2	−1.192	−0.046
Obs15	2	2	−1.192	−0.046
Obs22	2	2	−1.192	−0.046
Obs25	2	2	−1.192	−0.046
Obs49	2	2	−1.192	−0.046
Obs11	2	2	−0.952	−0.082
Obs21	2	2	−0.752	−0.073
Obs42	2	3	−0.452	0.182
Obs9	3	2	−1.192	−0.046
Obs16	3	2	−1.192	−0.046
Obs41	3	2	−1.192	−0.046
Obs47	3	2	−1.192	−0.046
Obs3	3	3	−0.452	0.182
Obs5	3	3	−0.452	0.182

(continued)

Table 5.7 (continued)

Observation	Prior	Posterior	F1	F2
Obs20	3	3	−0.452	0.182
Obs28	3	3	−0.452	0.182
Obs10	3	3	0.288	0.411
Obs14	3	3	0.288	0.411
Obs18	3	3	0.288	0.411
Obs8	3	1	0.741	0422
Obs43	3	1	0.741	0422
Obs23	3	1	1.028	0.640
Obs24	3	1	1.028	0.640
Obs19	3	1	1.768	0.869

scores of each observed variable, i.e. a news item on the two dimensional DA model.

Similar to EFA, the size of factor loadings indicates the relation between the observed variable and the conceptual scale constructed. A large positive score suggests that news items tend to exhibit textual features (sentiment analysis categories in this case) which have large positive loadings on that particular dimension. By contrast, a large negative score implies that news articles lack certain textual features or tend to exhibit reversed attributes of sentiment analysis categories which have large loadings on a given dimension. That is, factor scores, as shown in Table 5.7, enable the alignment and evaluation of strength of correlation between the observed variables, i.e. different news items and their associated sub-genres of translated environmental reporting with the predictor variables, i.e. the highlighted sentiment analysis categories, A1.2, A1.6, E4.2 and A5.2.

A glance at the loadings of news items on both dimensions leads to the discovery of the internal complexity or stratification of factor scores within each sub-genre of translated environmental news. To have a clear view of such differences, the news items belonging to each sub-genre were resorted in ascending order based on their computed factor scores on Dimension 1. This facilitated the isolation of news items with negative loadings from those with positive loadings on Dimension 1, which assumes an important discriminant function in the DA model. It should be noted that, given that there is no clear cut-off between sub-groups within each sub-genre, the identification of clusters with similar loadings on both dimensions necessarily involves subjective determination. However, this is common in many statistical analyses, especially with exploratory investigation in search of new theoretical models like the present study. Based on the resorting of news items within each sub-genre, a refined model of translated environmental news emerged, as shown below.

Sub-genre 1 (**Governance and Environment**)

Two sub-groups:

One with large negative loadings on Dimension 1 and small loadings on Dimension 2

Decreased use of evaluative language describing contentment

Decreased use of evaluative language describing abstraction and practicality

Neutralised use of abstract terms relating to appropriateness, suitability

Neutralised use of evaluative language expression false or true

One with small to large positive loadings on Dimension 1 and medium positive loadings on Dimension 2

Neutralised or increased of evaluative language describing contentment

Neutralised or increased of evaluative language describing abstraction and practicality

Moderate use of abstract terms relating to appropriateness, suitability

Moderate use of evaluative language expression false or true

Sub-genre 2 (**International Relations and Environment**)

Large negative loadings on Dimension 1 and small negative loadings on Dimension 2

Decreased use of evaluative language describing contentment

Decreased use of evaluative language describing abstraction and practicality

Neutralised use of abstract terms relating to appropriateness, suitability

Neutralised use of evaluative language expressing false or true

Sub-genre 3 (**Environmental Science**)

Two sub-groups:

One with large to small negative loadings on Dimension 1 and very small loadings on Dimension 2

Decreased or neutralised use of evaluative language describing contentment

Decreased or neutralised use of evaluative language describing abstraction and practicality

Neutralised use of abstract terms relating to appropriateness, suitability

Neutralised use of evaluative language expressing false or true

One with small to large positive loadings on Dimension 1 and medium loadings on Dimension 2

Increased or neutralised use of evaluative language describing contentment

Increased or neutralised use of evaluative language describing abstraction and practicality

Moderate use of abstract terms relating to appropriateness, suitability

Moderate use of evaluative language expressing false or true

The interpretation of the statistical results leads to important findings regarding distinctive framing strategies used in three types of environmental reporting published by BBC China. Four levels of the controlled use of evaluative and judgmental language are found in the three sub-genres of environmental news reporting on BBC China:

- Decrease (large negative loadings, above −1.0)
- Neutralisation (small negative or positive loadings, with the range of −0.4 and 0.4)
- Moderation (medium positive loadings: enhanced yet controlled use of evaluative language, within the range of 0.4 and 1.0)
- Increase (large positive loadings, above 1.0).

Sub-genre 1 (**Governance and Environment**)

Neutralised or moderate use of abstract terms relating to appropriateness, suitability

Neutralised or moderate use of evaluative language expressing false or true

Competing reporting style regarding the increased or decreased use of evaluative language describing contentment and abstraction/practicality

Sub-genre 2 (**International Relations and Environment**)

Decreased use of evaluative language describing contentment

Decreased use of evaluative language describing abstraction and practicality

Neutralised use of abstract terms relating to appropriateness, suitability

Neutralised use of evaluative language expressing false or true

Sub-genre 3 (**Environmental Science**)

Neutralised or moderate use of abstract terms relating to appropriateness, suitability

Neutralised or moderate use of evaluative language expressing false or true

Competing reporting style regarding the increased or decreased use of evaluative language describing contentment and abstraction/practicality

The corpus analysis revealed that there was more homogeneity within the sub-genre of international relations and environment, whereas competing styles of reporting existed for the sub-genres of governance efficiency and environmental science. The focus of debates and controversies in both cases was on the increased or decreased use of evaluative language describing contentment such as aggrieved, chuffed, content, dismay, disappointed, frustrated, humour, browned off, fed up, had enough of; and expressions conveying abstraction and practicality such as abstraction, hypothetical, metaphysics, notionally, practicalities, theorise or in theory. The tendency of purposely increasing evaluative language of contentment stands in contrast with other evaluative languages such as abstract terms relating to appropriateness, suitability and evaluative language expressing false or true, which have been consistently neutralised or reduced across the three sub-genres of environmental reporting.

The highly controlled use of attitudinal language reflects an important tradition of English-language journalism. Authorial "neutrality" is frequently seen to be a distinctive feature of the modern news report and one of the grounds on which journalists assert the objectivity of their writing. The increased use of evaluative language expressing contentment in translated environmental news, despite the close relation between the source and the target texts, could be due to various external factors including an effort made on the part of the translator and editor to re-conceptualise environmental issues and events so that they appear to have a more tangible impact on the readers' daily life; or as Hsieh (2008) observes, the level of subjectivity in Chinese news tends to increase when journalists reflect upon a reported event. It seems that Hsieh's finding regarding induced subjectivity in the reworking of reported news in Chinese journalism also holds in the current study. The result of the corpus analysis shows that Chinese translators and editors while working on the source English texts tended to modify the original texts for a culturally distinct audience. Translating original news materials represents a further process in the framing, adaptation and circulation of global environmental news to culturally diversified local audiences.

Another important finding emerged from the corpus analysis is the inconsistent reporting style of governance and environmental science regarding the presence of evaluative language expressing abstraction and practicality. An increased level of abstraction and theorisation in the reporting of environmental issues, particularly unobtrusive issues like climate change and global warming, requires specialised knowledge and a higher level of literacy from the audience. The co-existence of abstract and more concrete reporting styles seems to suggest that BBC China attempts to attract and appeal to readers with different educational backgrounds and political stances. A "balanced" account between abstraction and narration of issues which are seen as sensitive and controversial in China such as the efficiency or lack of efficiency of national governing bodies and authorities in dealing with environmental changes will necessarily add to the credibility, impartiality and depth of critical analysis to the news outlet.

A couple of important meta-frames or consistent textual patterns also emerged from the cross-genre analysis of translated environmental news. First, in all of the three sub-genres of translated environmental news, the use of abstract terms relating to appropriateness, suitability and evaluative language expressing false or true is neutralised (within the range of −0.4 and 0.4 in terms of factor scores) or maintained at a positive yet controlled level (within the range of 0.4 and 1.0). Secondly, a distinctive feature of the sub-genre of international relations and environment is the decreased use of evaluative language describing contentment, abstraction and practicality. The avoidance of judgmental language that evaluates the suitability and authenticity of elements of relevant news events, again, reflects the tradition of neutrality and objectivity of English-language journalism.

By combing the mixed findings of a tendency to modify and adapt original English news for culturally distinct audiences and a conspicuous adherence to the basic tenets and principles of English journalism, we could perhaps approach BBC China as a representative of localised digital media and news outlets that have been born out of the latest waves of globalisation. The sub-genres identified, their associated meta-frames and reporting strategies reveal the heterogeneity and dynamics of environmental reporting when the original news has been translated, adapted and delivered to specific cultural and language communities. What is worth paying attention to is the finding that mediated reporting, as represented by the translation of environmental news, tends to be more complex than original reporting in national contexts, since a variety of factors are involved in the production of mediated reporting. These include adherence to original reporting styles and underlying principles; and perhaps, more importantly, due consideration and accommodation of the expectations of the intended target readership.

5.6 Conclusion

The study reported in this chapter offers a first systematic investigation of mediated environmental news on BBC China. The investigation draws upon the sentiment annotation and analysis of translated environmental reporting, which represents an integral part of the heterogeneous sourcing system of the news outlet. The focus on translated news instead of original Chinese news is due to the higher level of credibility and reliability attributed to translation in the Chinese context (Tsien 1954; Hung 2002; Chan 2004). The prioritisation of translation over original texts written in the target language is, as Toury (2012) contends, the result of the power imbalance between exporting and importing cultural and social systems. With the advent of digital media and new communication technologies, globalisation has come to exhibit new patterns and trends which reflect the dynamics of ongoing cross-cultural and cross-lingual interactions. The focus of the current study, i.e. mediated or translated reporting of environmental issues, provides first-hand materials to investigate the process whereby translation takes the lead in promoting global consensus building beyond national boundaries.

In the case study presented here, BBC China was identified as the main news outlet from which relevant news data were collected, annotated and processed for quantitative and qualitative analyses. Differently from other major Western news outlets, which have been blocked and censured by the Chinese central government, BBC China is widely accessible from within the country for its perceived impartiality, neutrality and informative reporting. Our study investigated how issues like environment and development provide an important focus for the British news outlet to inform and mobilise its well-targeted Chinese-speaking audiences. Media framing analysis was used in an effort to examine the editorial strategies used by BBC China in the selection, translation and adaptation of original English environmental news.

Our corpus analysis shows that translated environmental news on BBC China adopts an approach to environmental reporting that differs from original environmental news. The latter tends to focus on a set of particular thematic issues in national contexts, for example, climate change and global warming in the USA; investigative reporting on industrial pollution and corporate responsibilities in China; natural disasters, nuclear energy pollution, management and changing life styles in Japan; international collaboration on environmental protection and economic development in the UK and north Europe. Our research detected three recurrent themes or parallel sub-genres of mediated environmental reporting published on BBC China. These are governance and environment, international relations and environment; and environmental science. Through the use of the Chinese version of the USAS sentiment analysis system developed by Lancaster University, UK, important subjective information was extracted from the three sub-genres of environmental reporting, particularly regarding the use of a variety of evaluative and attitudinal languages.

We found four distinct levels of the controlled use of evaluative languages in translated and mediated environmental reporting: decrease, neutralisation, moderation and increase. The delineation of these four levels is closely related to the loadings of individual news entry on two conceptual scales built by discriminant analysis (DA), which is widely used in behavioural and social sciences. At least four consistent patterns emerged from the cross-genre analysis of translated environmental news published on BBC China. First, we discovered the co-existence of increased or decreased use of evaluative language describing contentment as well as expressions conveying abstraction and practicality in the sub-genres of governance efficiency and environmental science. Second, we found an inconsistent reporting style of governance and environmental science regarding the presence of evaluative language expressing abstraction and practical value In contrast to the inconsistencies in the use of some subsets of attitudinal languages, regularities do seem to exist across the three sub-genres. Any consistent patterns in the distribution of attributes and features of evaluative and subjective languages were then approached from the perspective of meta-frames of reporting which represent the core of media frame analysis.

There are two important meta-frames which are characteristic of mediated environmental reporting on BBC China, i.e. the neutralisation of abstract terms relating to appropriateness and suitability, and evaluative languages which express what is true or false. Across the three sub-genres, individual news texts have loadings within the range of −0.4 and 0.4 on the conceptual scale which hosts these two subsets of evaluative languages. Second, with regard to the sub-genre of international relations and environment, which has the highest accuracy score in DA, the decreased use (with large negative loadings above −1.0) of evaluative languages describing contentment or disapproval, abstraction and/or practicality is noticeable and consistent.

The corpus findings concerning competing reporting styles and the existence of consistent meta-frames across the three sub-genres of environmental news have two implications. First, there is an explicit adherence to basic tenets, principles and ethos of English-speaking journalism such as emotional distance, neutrality and balanced reporting. This is supported by the overall decreased or neutralized use of attitudinal languages expressing appropriateness, suitability and true or false. Given the controversy and sensitivity of some environmental issues that border on the (in) effectiveness and (lack of) long-term viability of some governmental policies, especially corruption of local governments in China, this is deemed to be a sensitive reporting strategy to secure the presence of the British news outlet in China.

Second, mediated or re-framed environmental news published on BBC China do not represent a literal translation of original English news. Competing reporting styles as reflected in the inconsistent use of certain attitudinal languages do seem to exist, especially as regards the sub-genre of governance and environment. The corpus analysis shows that while some news texts explicitly increased the use of evaluative languages describing contentment or disapproval, other randomly selected texts exhibited a reversed pattern. This finding supports and fits in well with what Tong (2014) calls the rise of critical reflective discourse in China, which has gained in popularity in recent years with the ongoing privatisation with some shares of the Chinese media markets.

These two underlying patterns of mediated or translated and adapted environmental reporting, i.e. adherence to essential principles and tenets of English-speaking journalism and tendencies to align with emerging patterns of environmental reporting in local contexts and to accommodate expectations of the targeted audiences, reveal important strategies developed by BBC China. These important strategies on the one hand, help the British news outlet to secure and enhance its presence in China; and on the other hand, play an instrumental role in promoting the construction of global consensus around issues which require collaborative efforts and shared understanding beyond national boundaries as well as language and cultural differences.

5.7 Future Research

As a first investigation of mediated environmental reporting published on the localised regional version of a major English news outlet, BBC China, this study has explored the validity and productivity of the automatic sentiment analysis system (USAS) developed by the Centre for Computer Corpus Research on Language of the University of Lancaster, UK. While the USAS system has been well tested in a number of studies (Gabrielatos and McEnery 2005; McIntyre and Walker 2010; Shapero and Blackwell 2012; Woodworth et al. 2012; Potts and Baker 2012), its Chinese version has not been explored sufficiently in applied linguistics. In the computational processing, especially the automatic annotation stage of the Chinese translated environmental news data base, some technical issues which reflect the developing nature of CH-USAS began to emerge. First, a large number of Chinese character words were labelled as Z99. This is the last item in the USAS category system, which includes misspellings or words that have not been included in the lexicon as yet. In the post-editing process, which involved human annotation and checking, we found that an important portion of character words that were labelled as Z99 are subject-specific terms and expressions which are closely related to environmental issues, sustainable development; Chinese idiomatic expressions; and more importantly, new terms and words which have been created in the translation process.

This technical issue has to do with the lexical resources, printed or digital, that were used in the development of the Chinese version of USAS. Environment and development represents a rapidly growing new area of research. Many new terms and expressions that make up the core lexicon of the field have not been collected into general purpose dictionaries and digital term bases like those used in the development of CH-USAS. For well over a century, translation has provided an important source for the scientific modernisation and the assimilation of new idea sets and concepts imported from the West into the native knowledge body of China (Wright 2000; Ji 2013b). The latest public debates, scholarly interests and scientific pursuits in environment and development in China is no exception to this tradition. As the analysis of the translation data base of environmental news shows, an increasing number of new character words have been created from original English texts, for example, 臭氧层 (ozonosphere); 氯氟化 (chlorofluorination); 可再生能源 (renewable energy); 低碳生活 (low-carbon living); 海洋酸化 (ocean acidification); 核污染 (nuclear pollution); 温室效应 (greenhouse effect); 厄尔尼诺 (El Niño Southern Oscillation), to name but just a few.

Once incorporated into the current version of the CH-USAS system, these new terms and expressions could effectively increase the overall accuracy of the automatic sentiment annotation, as lexical under-specificity represents a bottleneck in the current subjective information mining system, as far as the Chinese language is concerned. As noted in the present study, given the cost and labour involved in the manual revision process, which rectified multiple labelling and mislabelling by the CH-USAS alongside lexical under-specification, the size of the translation data base was limited. This had an inevitable impact on the subsequent quantitative and qualitative analysis. However, we believe that with the improvement of these technical issues, the USAS system will facilitate large-scale systematic investigation of environmental reporting by extracting subjective information and different varieties of evaluative and attitudinal languages from news texts in sufficient quantities and with high levels of accuracy and precision.

An important question that remains to be answered is how the meta-frames detected in the quantitative analysis of translated environmental news apply to environmental news from other sources and in the source language, i.e. British English and the target language, i.e. Chinese. How do translated environmental news serve as a mediated news genre to promote understanding between different readerships and global consensus building around issues like environmental protection, international collaboration on climate change and global warming? Are the textual patterns, meta-frames and related reporting strategies discovered in the case study of BBC China applicable in other localised and mediated news websites? Would reversed translation direction, for example, translation of original Chinese news into English exhibit similar or contradictory patterns to those found in Western news outlets? To address these issues and questions, it is necessary to construct large-scale comparable data bases or corpora in the source and the target languages to verify, modify and expand the analytical models used in this study.

References

Alin, A. 2010 Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(3): 370–374.

Andersen, R.S. 2007. *The WTO Agenda and the Media Agenda: Villains, Victims and the International Press Coverage of the Doha Development Agenda Negotiations*. Department of Political Science, University of Copenhagen.

Baker, M. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target*, *7*(2): 223–243.

Bassnett, S. 1998. The translation turn in cultural studies. *Constructing Cultures: Essays on Literary Translation*, 123–140.

Bernardini, S. 2005. Reviving old ideas: Parallell and comparable analysis in translation studies-with an example from translation stylistics. *Gothenburg Studies in English* 90: 5–18.

Bosseaux, C. 2004. Point of view in translation: A corpus-based study of French translations of Virginia Woolf's To The Lighthouse. *Across Languages and Cultures* 5(1): 107–122.

Boykoff, M.T. 2007. From convergence to contention: United States mass media representations of anthropogenic climate change science. *Transactions of the Institute of British Geographers* 32(4): 477–489.

Boykoff, M.T. 2008. The cultural politics of climate change discourse in UK tabloids. *Political Geography* 27(5): 549–569.

Boykoff, M.T., and J.M. Boykoff. 2004. Balance as bias: global warming and the US prestige press. *Global environmental change*, *14*(2):.125–136.

Carvalho, A. 2007. Ideological cultures and media discourses on scientific knowledge: Re-reading news on climate change. *Public Understanding of Science* 16(2): 223–243.

Carvalho, A., and J. Burgess. 2005. Cultural circuits of climate change in UK broadsheet newspapers, 1985–2003. *Risk analysis*, *25*(6): 1457–1469.

Chan, L.T.H. 2004. *Twentieth-century Chinese translation theory: Modes, issues and debates*, vol. 51. Amsterdam: John Benjamins Publishing.

David, C.C., J.M. Atun., E. Fille and C. Monterola. 2011. Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures* 5(4): 329–351.

Deignan, A., and L. Potter. 2004. A corpus study of metaphors and metonyms in English and Italian. *Journal of Pragmatics* 36(7): 1231–1252.

Dirikx, A. and D. Gelders. 2010. To frame is to explain: A deductive frame-analysis of Dutch and French climate change coverage during the annual UN Conferences of the Parties. *Public Understanding of Science*, *19*(6): 732–742.

Dorothy, K. 2001. *Lexis and creativity in translation: A corpus Based approach*. London: Routledge.

Franklin, B. 2014. The future of journalism in an age of digital media and economic uncertainty. *Journalism Practice* 8(5): 469–487.

Gabrielatos, C., and T. McEnery. 2005. Epistemic modality in MA dissertations. In Fuertes Olivera, P.A. (ed.) Lengua y Sociedad: Investigaciones recientes en lingüística aplicada. Lingüística y Filología no. 61. Valladolid: Universidad de Valladolid, pp. 311–331.

Gideon, T. 1995. *Descriptive translation studies and beyond*. Amsterdam and Philadelphia: John Benjamins.

Goethals, P. 2007. Corpus-driven hypothesis generation in translation studies, contrastive linguistics and text linguistics: A case study of demonstratives in Spanish and Dutch parallel texts. *Belgian Journal of Linguistics* 21(1): 87–103.

Granger, S. 2003. The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. Corpus-based Approaches to Contrastive Linguistics and Translation Studies, 17–29. Amsterdam: Rodopi.

Hope, M. 2010. Frame Analysis as a Discourse-Method: Framing 'climate change politics. In Proceedings of Post-Graduate Conference on Discourse Analysis, Bristol, UK.

Hsieh, C.L. 2008. Evidentiality in Chinese newspaper reports: subjectivity/objectivity as a factor. *Discourse Studies* 10(2): 205–229.

Huang, S., and D. Graff. 2002. *Chinese and English Translation Lexicon, LDC2002L27*. Philadelphia: Linguistic Data Consortium.

Hung, E. 2002. Translation and English in twentieth–century China. *World Englishes* 21(2): 325–335.

Janetos, A.C., E. Malone., E. Mastrangelo., K. Hardee., and A. de Bremond. 2012. Linking climate change and development goals: Framing, integrating, and measuring. *Climate and Development* 4(2): 141–156.

Ji, M. 2013a. *Exploratory statistical techniques for the study of literary translation*. Lüdenscheid: RAM-Verlag.

Ji, M. 2013b. *A corpus-based study of historical Chinese Lexis*. Tokyo: Waseda University Press.

Malmkjær, K. 2004. Translational stylistics: Dulcken's translations of Hans Christian Andersen. *Language and Literature* 13(1): 13–24.

Mauranen, A., and P. Kujamäki (eds.). 2004. Translation universals: do they exist? John Benjamins Publishing.

McArthur, T. 1986. Longman lexicon of contemporary english, Pearson P T.

McIntyre, D., and B. Walker. 2010. How can corpora be used to explore the language of poetry and drama? In *The Routledge handbook of corpus linguistics*, ed. McCarthy, M. and A. O'Keefe, 516–530. Routledge.

Oakes, M. and M. Ji. 2012. *Quantitative Methods in Corpus-Based Translation Studies*, Amsterdam: John Benjamins.

Olohan, M., and M. Baker. 2000. Reporting that in translated English. Evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1(2): 141–158.

Pastor, G. C., Mitkov, R., Afzal, N., and V. Pekar. 2008. Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. In *Proceedings of the 8th AMTA Conference*, 75–81.

Potts, A., and P. Baker. 2012. Does sentiment tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics* 17(3): 295–324.

Puurtinen, T. 2003. Explicitating and implicitating source text ideology. *Across Languages and Cultures*, *4*(1): 53–62.

Rooke, R. 2013. *European media in the digital age: Analysis and approaches*. London: Routledge.

Scarpa, F. 1999. Corpus evidence of the translation of genre-specific structures. *Textus* 12(2): 1000–1018.

Schäffner, C. 2004. Metaphor and translation: Some implications of a cognitive approach. *Journal of Pragmatics* 36(7): 1253–1269.

Shapero, J.J., and S.A. Blackwell. 2012. 'There are letters for you all on the sideboard': What can linguists learn from multiple suicide-note writers? In *Proceedings of The International Association of Forensic Linguists' Tenth Biennial Conference*, ed. Samuel Tomblin, Nicci MacLeod, Rui Sousa-Silva, and Malcolm Coulthard, 225–244. Centre for Forensic Linguistics: Aston University, U.K.

Shuttleworth, M. 2011. Translational behaviour at the frontiers of scientific knowledge: A multilingual investigation into popular science metaphor in translation. *The Translator* 17(2): 301–323.

Snow, D.A. and R.D. Benford. 1988. Ideology, frame resonance, and participant mobilization. *International social movement research*, *1*(1): 197–217.

Stefanowitsch, A., and S.T. Gries. (eds.), 2007. Corpus-based approaches to metaphor and metonymy (Vol. 171). Walter de Gruyter.

Tabachnick, B.G., and L.S. Fidell. and Osterlind, S.J., 2007. *Using multivariate statistics*. Boston: Pearson Education.

Tong, J. 2014. Environmental risks in newspaper coverage: A framing analysis of investigative reports on environmental problems in 10 Chinese newspapers. *Environmental Communication: A Journal of Nature and Culture* 8(3): 345–367.

Toury, G. 2012. Descriptive Translation Studies and beyond: Revised edition (Vol. 100). John Benjamins Publishing.

Trosborg, A. (ed.). 1997. *Text typology and translation*, vol. 26. John Benjamins Publishing.

Tsien, T.H. 1954. Western impact on China through translation. *The Journal of Asian Studies* 13(3): 305–327.

Woodworth, M., J. Hancock., S. Porter., R. Hare., M. Logan., O'Toole, M.E. and S. Smith., 2012. The language of psychopaths: New findings and implications for law enforcement. *FBI Law Enforcement Bulletin*. http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/july-2012/the-language-of-psychopaths.

Wright, D. 2000. *Translating Science: The transmission of western chemistry into late imperial China, 1840–1900*, vol. 48. Leiden: Brill.