

Normalization in Translation

Normalization in Translation:
Corpus-based Diachronic Research
into Twentieth-century English–Chinese
Fictional Translation

By

Yun Xia

CAMBRIDGE
SCHOLARS

P U B L I S H I N G

Normalization in Translation:
Corpus-based Diachronic Research into Twentieth-century English–Chinese Fictional Translation,
by Yun Xia

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2014 by Yun Xia

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system,
or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or
otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-6037-9, ISBN (13): 978-1-4438-6037-6

TABLE OF CONTENTS

List of Tables	vii
List of Figures.....	viii
Acknowledgements	ix
List of Abbreviations	x
Chapter One.....	1
Introduction	
Chapter Two	12
Corpus Methodology in Translation Studies	
Chapter Three	36
Creating and Using the Diachronic Corpus	
Chapter Four	54
Lexical Normalization	
Chapter Five	75
Normalization of Collocation	
Chapter Six	96
Syntactic Normalization	
Chapter Seven.....	126
Contextualization of Textual Norms	
Chapter Eight.....	144
Conclusions	
Appendix A	151
Appendix B.....	154

Appendix C..... 155

Appendix D 157

Notes..... 159

Bibliography 162

Index 179

LIST OF TABLES

- Table 3-1. General features of the two translated subcorpora
Table 3-2. Overall size of each subcorpus
Table 4-1. POS distribution in the four Chinese subcorpora
Table 4-2. Distribution of nouns in TCT1 and OCT1
Table 4-3. Distribution of nouns in TCT2 and OCT2
Table 4-4. Frequency difference of pronouns in translated vs. original texts
Table 4-5. Frequency data of personal pronouns
Table 4-6. Overlapping rate of high-frequency words
Table 4-7. Compositionality of five quasi-affixes
Table 5-1. Frequency of selected nodes
Table 5-2. Collocates of selected nodes
Table 5-3. Lexical collocates of the node 话 in TCT2
Table 5-4. Frequency percentages of the top collocates of each node in TCT1 and OCT1
Table 5-5. Frequency percentages of the top collocates of each node in TCT2 and OCT2
Table 5-6. Repetition rate of the *a* (的/底) *n* collocation
Table 5-7. Frequency of major translation methods
Table 6-1. Overall frequency data of connectives
Table 6-2. Categories of TDCs in TCT1 and TCT2
Table 6-3. Overlapping rate of connectives in translated vs. non-translated texts
Table 6-4. Frequency data of prepositions
Table 6-5. Top 6 prepositions in each subcorpus
Table 6-6. Bilingual concordance results of 可是
Table 6-7. Frequency data of Chinese passive markers
Table 6-8. Long and short passives in different subcorpora
Table 6-9. Long and short agents in long passives
Table 6-10. Mean sentence length and sentence segment length
Table 6-11. Number of long SS (15-20 words in length)
Table 6-12. Distribution of major types of headword

LIST OF FIGURES

- Figure 2-1 Flowchart of a corpus-assisted translation study
- Figure 3-1. Composition of DPCECF
- Figure 3-2. Model for a diachronic study of normalization
- Figure 3-3. Flow chart of corpus construction
- Figure 4-1. A broken line graph of POS distribution
- Figure 4-2. Frequency of common nouns and proper nouns in source texts and translations
- Figure 4-3. Distribution of different types of pronouns in each subcorpus
- Figure 4-4. Frequency proportion of high-frequency words
- Figure 4-5. Proportion of mono- and poly- syllabic words
- Figure 4-6. Proportion of monosyllabic words in translated texts
- Figure 4-7. Normalized frequencies of 5 affixes in TCT1 and OCT1
- Figure 4-8. Normalized frequencies of 5 affixes in TCT2 and OCT2
- Figure 6-1. Frequency of connectives in each translated text pair
- Figure 6-2. Frequency percentage of long passives with *bei* and *gei*
- Figure 6-3. Distribution of agent length in long passives
- Figure 6-4. Semantic features of syntactic passive constructions

ACKNOWLEDGEMENTS

My special thanks first go to the Humanities and Social Sciences Foundation of the Ministry of Education of China for supporting my research project “A Corpus-based diachronic Study of Normalization in English–Chinese Translated Fiction” (grant reference 10YJC740108). I am also grateful to grants from China National Foundation of Social Sciences (grant reference 10BYY008), National Grants for Social Sciences Major Projects (grant reference 10ZD&127), the China Scholarship Council (CSC) and the Doctoral Research Fund of Qufu Normal University for being able to work intensively on a research project of my choice.

The research conducted for this book would not have been possible without the professional guidance from Professor Defeng Li, my Ph.D. supervisor, from the School of Oriental and African Studies, University of London. I am also grateful to Professor Hongwu Qin for his precious advice on my corpus construction and the use of corpus tools for corpus analysis. I should also thank Doctor Richard Xiao in Edge Hill University, U. K. for kindly sharing the invaluable corpus resource, LCMC, which is of great value to this study.

—Yun Xia
April 2014

LIST OF ABBREVIATIONS

DPCECF	Diachronic Parallel Corpus of English–Chinese Fictions
OCT1	Original Chinese texts of 1930s and 40s
OCT2	Original Chinese texts of published in 1989–1993
S	Sentence
SS	Sentence segment
TCT1	Translational Chinese texts of 1930s and 40s
TCT2	Translational Chinese texts published since 1988
TDC	Translation-dominant conjunctions
TOW	Translation-only words

CHAPTER ONE

INTRODUCTION

1.1 Why a Diachronic Corpus-based Approach?

The use of electronic corpora—large bodies of authentic texts in machine-readable format—within translation studies has become common in terminology compilation and machine translation. In descriptive translation studies, there has also been rapid development of the corpus-based approach in recent years. Corpus analysis allows us to see phenomena that previously remained obscure because of the limitations of our vantage points, and is applied to the study of translation product and process. As a research method, it may therefore be employed in researching questions and issues in conjunction with a range of theoretical frameworks, assumptions, tools and concepts. Corpus-based translation studies (CTS) focuses on both the process and the product of translation, and takes into account the smallest details of the text chosen by the individual translator, as well as the largest cultural patterns both internal and external to the text (Tymoczko 1998). Research in the area covers a variety of studies, ranging from descriptive studies on norms/universals of translation and translator style, to applied studies like translator training. Maeve Olohan's *Introducing Corpora in Translation Studies* (2004) provides an excellent overview of research conducted in the area.

At the core of descriptive CTS since the 1990s is the idea that there are certain tendencies that are indicative of “the nature of translated text in general and the nature of the process of translation itself” (Baker 1995: 236) and that become manifest in recurrent patterns in the linguistic make-up of translated texts. Studying what distinctive features translated texts exhibit and how they differ from original, non-translated texts written by native writers has been a topic of long-standing interest in translation studies. Initial research goes back to Toury (1995) who put forward the “law of growing standardization” and the “law of interference”, and it was Baker (1993, 1996) who first formulated some of the so-called universal features of translation. Research on “translation universals” is conducted typically by comparing corpora of translated and non-translated texts in

the target language. The pursuit of universals shares much in common with the interest in laws of translation promoted in Descriptive Translation Studies (DTS), behind which lies the presupposition of Western rationalism that science should be in the business of discovering natural laws and that “scientific” results have more value than others (Tymoczko 1998: 2).

One such universal tendency most intensively studied in the literature is normalization, defined by Baker (1996) as the “tendency to conform to patterns and practices which are typical of the target language, even to the point of exaggerating them”. While evidence has been uncovered to support the hypothesis of normalization, counter-evidence has also been found showing the coexistence of normalized and non-normalized features in translation. “Different times and cultures may well conceptualize the notion of ‘translation’ in very different ways” (Chesterman 2006), and consequently, the notion of normalization as a “universal” feature became a target of debate. It remains questionable whether normalization is to be reckoned as a “translation universal” which characterizes translation as a “third code” (see Frawley 1984: 168), or subject to socio-cultural variation. Does it apply to different historical conditions? To what extent do normalized/non-normalized features of translation influence native language development? A diachronic study of normalization in translations of different historical periods, therefore, will prove useful, and should hopefully shed light on the question of whether such a feature is universal-induced or norm-induced. Such a study necessarily requires taking historical-cultural factors into account, that is, an assessment of the impact and interplay of all possible conditions constraining the translational behaviors is necessary. However, there has been a lack of studies on regularities in translational behavior from the diachronic perspective, and previous studies have focused mainly on the differences between general tendencies in translated and non-translated texts from a synchronic perspective, more or less aimed at testing the underlying intuition that translations seem to exhibit normalized or more conventional linguistic features regardless of linguistic differences and socio-cultural variation. Furthermore, recent research tends to rely heavily on electronic corpora, which are unavoidably decontextualized. Tymoczko (1998: 658) objects to the notion of laws or universals, and argues against a blinkered perspective within corpus-based studies that does not see the past, and other cultures, but that is instead “locked into the translation norms of the present, [...] presupposing such norms in the construction of corpora”. The current investigation, therefore, has been designed to expand the scope of inquiry by exploring normalization as a norm-induced feature from a diachronic perspective.

Specifically, there are a number of purposes to undertaking empirical studies on normalization from a diachronic perspective. The first objective is to explore variation of normalization across times, which might shed light on the nature of the notion: Is normalization a “universal” feature of the translation process or does it reflect the changeable translation norm which is in turn subject to changes in the socio-cultural environment? For this purpose, two collections of Chinese fictional texts translated in two time spans from the same set of English source texts are constructed, together with non-translated Chinese texts of the same genre from the same time spans as the control corpora. Generally speaking, normalization has rarely been systematically investigated, and most previous corpus-based studies have been restricted to single aspects of its linguistic manifestation. A case in point is Kenny (2001), one of the most important large-scale research projects on normalization to date, who explores this issue on the lexical level. The present study attempts to expand the scope of inquiry into the nature of normalization, by a multi-level analysis of manifestations of normalization on different linguistic levels in English–Chinese translated texts of different historical periods, involving words, word combinations, and syntactic features, and by exploring whether or not changes in such manifestations take place across time spans. Such work is based on the assumption that translators do tend to “normalize” their translations, and that normalization could be discerned in the surface make-up of translations as products. The term “normalization” in the present research is taken as a relative notion, a dimension that might be used to measure the socio-culturally variable translation norms in different historical periods.

The second objective in undertaking this study is to highlight the social and cultural significance of translation activities of different historical backgrounds, by relying on first-hand observation of corpus research findings, and by exploring the socio-cultural factors in different historical contexts in China which might account for the observed similarities and differences in translation behavior across times. Since the late nineteenth century, translation—English–Chinese translation in particular—has played a crucial role in the development of modern Chinese, and the influence of translation on native Chinese was deeper in the twentieth century. Researchers who have taken a special interest in research into the influence of translation on Chinese have found an “Europeanization” tendency in modern Chinese at both lexical and syntactic levels (Wang 1943, Xie 2001, Wang 2002). According to Wang (1943), Chinese has undergone marked changes since the “May 4th Movement” (May 4, 1919) due to Western language influence, mainly manifested by changes in word

usage, parts of speech, frequency of particular sentence structures, formalization of syntax, and position of subordinate clauses in complex sentences. Wang (2002) discusses the impact of translation on Chinese syntax as manifested by increase in passive voice and greater complexity and variability of sentence structures, and further points out that although influence from Western languages on Mandarin Chinese, largely through translation, is not such as to produce Europeanized Chinese, it is nevertheless a key factor in promoting language change, especially in terms of lexis and syntax.

Though previous studies have taken into account translation as a factor in triggering native language changes and have emphasized the importance of taking into account the interplay of textual features and extratextual factors, insufficient efforts have been made toward a systematic corpus-based macro analysis of the interaction between translation and native Chinese and the contextualization of corpus findings. A target-oriented exploration of historical factors in this study should hopefully offer new insights into the nature of translation as a socio-cultural product, and the cultural significance that translation plays in particular historical periods of modern China. Through such analyses, the research also attempts to prove that studies on translation activities will be imperfect and partial if it is only based on generalizations inferred from linguistic regularities observed without considering the restraints and needs of the target culture.

Third, the present study aims to refine the methodology used in previous research, by relying on a combination of synchronic and diachronic perspectives, and by combining monolingual comparable corpus and parallel corpus resources. Previous corpus-based studies have provided a methodological model for investigating recurrent features of translation and their underlying norms/universals. However, corpus-based diachronic study has rarely been conducted on regularities of translational behavior, and most previous studies have relied simply on monolingual corpora or parallel ones for their research on features of translation. The present study has built upon the methods used in the previous studies, applying them to materials of different time spans from a diachronic perspective. Such a new perspective will not only offer insights into the nature of translation activities, but surely bear some methodological implications. It is clearly related to the contextualization of translation and a tracing of the link between “regularities of actual behavior” (Toury 1995: 265) and “interests and stakes being pursued” (Hermans 2000: 12–13), and may help to promote the integration of cultural studies of translation with the systematic methods of corpus linguistics. Furthermore, by constructing the diachronic

corpus which includes a composite set of monolingual and parallel resources, this study aims to show how different resources all have their role to play in the research programs of corpus-based translation studies. The types of searching techniques and measures in the present study have also been further extended, including word length, overlapping rate of high-frequency words, affixation, etc., some of which to the best of our knowledge have not previously been used in corpus-based translation studies. It is hoped that the study can lead to a better understanding of the nature of normalization and norms of translation, and that the study can pave the way for subsequent research.

1.2 Normalization in Translation

1.2.1 Defining Normalization

The hypothesis of normalization, originally proposed as “conservatism” by Baker (1996: 183), which is also called “conventionalization” in the literature (e.g., Mauranen 2007), states that the translator tends to “conform to patterns and practices typical of the target language, even to the point of exaggerating them”. Kenny (1998:1–6) proposes “sanitization” as a corollary of lexical normalization, and finds that at points where source text collocations evoke difficult-to-read, controversial, ironic, or unpleasant (semantic) associations, their corresponding target texts may tend to switch toward more palatable imagery and metaphor. Williams (2005: 8) proposes a general hypothesis of normalization, and observes that “the written texts produced by a population of translators will conform more closely to the norms prevailing for written texts in the language of translation, and this will result in observably fewer instances of atypical or “non-conformist” usage in translated texts [...] Specifically, we will find fewer examples of unattested words and phrases in translated corpora”.

This assumption of normalization in translation could also be discerned in DTS scholars such as Even-Zohar (1990: 48–49), who claims that translated literature is always “modeled according to norms already conventionally established” and is in fact a “major factor of conservatism” in the literary polysystem of a culture. In the literature on translation norms, we have noted a general assumption that translators are conformist, following language norms more closely than authors do, and that where the rules of language are concerned, a greater degree of conformity is seen in translators than in other writers (see also Williams 2005: 20). As a result, translational language appears to be used in a more conventional or normalized way than the target language. This means that though

translation, just like writing, is a type of social activity generally subject to the same set of language norms as original writings are, the translator seems to be under greater pressure to produce “acceptable” (or marketable) writing than the original writer is.

Based on previous analyses of normalization, we might define the notion as referring to the tendency of translators to conform to the conventionally established and standard practices typical of the target language, which would in turn result in relatively higher conservativeness and repetitiveness, and a lower diversity of language use in translations as well.

Typical manifestations of normalization might include overuse of clichés or typical grammatical structures of the target language, adapting punctuation to the typical usage of the target language, and the treatment of the different dialects used by certain characters in dialogues in the source texts. (Xiao, He and Yue 2010: 187). Baker (1996: 183) says that normalization is manifest in grammaticality, typical punctuation and collocational patterns. Malmkjaer (1997) and Øverås (1998) maintain that in literary translation, the language is more likely to be lexically conventional, with fewer unusual word combinations. Munday (1998) believes that (literary) translated texts are more likely to have conventional word order.

1.2.2 Normalization, Norms and Universals of Translation

Norm is a key concept in DTS research. Translation in DTS is seen as a communicative act that constitutes a form of social behavior, subject to socio-cultural constraints specific to a culture, society and time which are imposed by more-or-less binding intersubjective norms. For Toury, norms are regarded as “the translation of general values or ideas shared by a group—as to what is conventionally right and wrong, adequate and inadequate—into performance instructions appropriate for and applicable to particular situations, specifying what is prescribed and forbidden, as well as what is tolerated and permitted in a certain behavioral dimension” (Toury 1995: 54–55). Hermans (1996) distinguishes between conventions, norms, and rules. Norms, like rules and conventions, have a socially regulatory function, and help to bring about the coordination required for continued coexistence with other people. Conventions are a matter of “reciprocal expectations”, and they are implicit norms at best, whereas rules are strong norms, usually institutionalized and posited by an identifiable authority (Hermans 1996: 29–32). According to Hermans (1999b: 80), the term “norm” refers to both a regularity in behavior, i.e., a

recurring pattern, and to the underlying mechanism which accounts for this regularity. The mechanism is a psychological and social entity.

Chesterman (1997) draws on the notion of “memes” from sociology, which is simply an idea that spreads, and translation norms are seen as an essential set of professional translation memes. “Each time a translator conforms to a norm, and thus spreads the meme, the norm becomes stronger. But each time someone breaks the norm, it becomes a bit weaker” (Chesterman 2000: 7). Despite the differences in defining the notion, it is generally accepted that norms are historical entities, and hence subject to change as they adjust to changing circumstances. Some norms, though, are more robust and durable than others. Researchers in DTS have also noted the complexity of norms. Toury, for example, observes that norms interact in complex ways and that they are inherently variable and unstable (Toury 1995: 59–62). Hermans (1996: 39) also notes that when studying translations, we may expect to find “a variety of competing, conflicting and overlapping norms and models which pertain to a whole array of other social domains”.

Translation norms have been analyzed and classified in many ways (see, e.g., Toury 1995, Chesterman 1997, Schäffner 1998, Hermans 1999a). Some norms govern the relation between the source text and the target text; others relate to the form of the target text, and target-language conventions of style, etc. Toury, for example, distinguishes three specific norms that may operate in any translation activity: (1) preliminary norms, which relate to translation policy on what texts to be selected for translation, and to the overall translation strategy; (2) initial norms, which inform the basic choice to adhere to the source text or the target culture; and (3) operational norms, which govern the very existence of target text material and the actual linguistic formulation of the target text. Chesterman (1993, 1997) differentiates between expectancy norms and professional norms. The former refers to what the target language community expects a translation to look like “regarding grammaticality, acceptability, appropriateness, style, textuality, preferred conventions of form or discourse and the like” (Chesterman 1993: 17). The latter governs the accepted methods and strategies of the translation process, which can be further divided into accountability norms, communication norms and relation norms.

Though classified in different ways, the notion of “norms” discussed by Toury and Chesterman basically involves two aspects of translation norms: one relating to the external socio-cultural context (reader expectancy about the correctness and suitability of translation, power relations, translation policy, etc.) that governs the whole translation

activities, which might be called “extratextual norms”; the other relating to textual norms that govern the actual linguistic operation of the translator in the translation process, and could be discerned in the regularities of translational behavior or recurrent features of translation. We could say that textual norms are linguistic manifestations of the translator’s behavior and strategies at the surface structure of texts and are subject to the constraints of extratextual norms.

While norms are socially and culturally determined and vary across time and space, translation universals are generally described as a product of constraints inherent in the translation process. According to the hypothesis of “Translation Universals”, any translated language variety, regardless of the source and target languages, might share characteristic features typical of translation “as a mediated communicative event” (Baker, 1993). That is, they are not the result of interference from either source or target language and would not vary across cultures. Unlike norms, universals of translation are relatable to cognitive factors rather than social ones (Kenny 2001: 53). There are researchers who argue against the notion of translation universals. Toury (2004: 29), for example, feels that so-called “universals” of translation, such as explicitation, cannot be understood to cover every act of translation since no features of translation are ever “universal” unless they are so general and bland as to be of little use (e.g., “translation involves shifts”). Instead, Toury prefers the term “law” to “universal” because “this notion has the possibility of exception built into it [and] because it should always be possible to explain away (seeming) exceptions to a law with the help of another law, operating on another level”. Tymoczko (1998) also believes that it is not possible to formulate universal, or general, laws of translation, not least because of changing views, across cultures and through time, of the concept of translation. This means we cannot talk about universals of translation or universal laws of translation if we cannot account for all translation.

Normalization was proposed earlier as one of the possible “universals” of translation by some scholars. This feature seems to have much in common with Toury’s (1995) “law of growing standardization” as opposed to the “law of interference”. However, Toury’s laws are probabilistic in the sense that their operation is conditioned by socio-cultural factors such as the status of translation within the target culture. Toury suggests that the more peripheral translation is, the more it will accommodate itself to established models and repertoires in the target polysystem (1995: 271). Similarly, rather than treat the notion as a universal feature of translation, Kenny (2001: 67) argues that normalization may be more easily explained in terms of socio-cultural, or even economic constraints, and that in

cultures where the norm biases translations towards target language and culture acceptability, translated texts that deviate from such target acceptability may run the risk of being ignored, criticized or ultimately rejected by their intended audiences, and so involve higher financial risks for publishers. This implies that the feature of normalization might not be uniform in translations of different cultures, historical periods and text genres. It might be subject to changes in social and historical circumstances.

There are researchers who have come to notice the controversial nature of the concept (Xiao *et al.* 2010) with evidence being found of the existence of innovative translation. Toury (1995: 31) observes that one of translation's universal characteristics is its variety, or "variability". According to Toury (1995: 208), it is a "well-documented fact that in translations, linguistic forms and structures often occur which are rarely, or perhaps even never encountered in utterances originally composed in the target language". Kenny's (1998, 2000a, 2000b and 2001) series of studies of literary translations also reveal a coexistence of normalizing and non-normalizing shifts in translation. Even Baker (1998) herself admits that the consistent language choice by the translator does not necessarily mean that these choices are uniformly and without exception conservative. Actually, Baker (1998: 5) points out that non-normalized translation choices are not necessarily erroneous, and that they are not insignificant to translation studies (see also Williams 2005: 20).

We can, therefore, assume that normalization, as regularities of actual translation behavior, is induced by translation norms rather than universals, hence subject to changes in socio-cultural factors. But rather than equating normalization with norms themselves, the present study takes normalization as a potential surface-structure realization of textual norms, following Toury's view of translational norms in distinguishing between norms themselves and the realization of such norms.

Needless to say, whatever regularities are observed, they themselves are not the norms. They are only external evidence of the latter's activity, from which the norms themselves are still to be extracted, whether by scholars wishing to get to the bottom of a norm-governed behavior or by persons wishing to be accepted in the group and hence needing to undergo socialization. (Toury 1999: 15)

One reason for distinguishing between the two is that practice has found it "difficult to associate unambiguously many observable patterns with a unique underlying norm" (Kenny 2001: 51). Observations that norms are psychological and social entities (Hermans 1996: 26, Chesterman

1997), and that they are of an abstract nature (Kenny 2001: 50) also suggest that the two ought to be kept apart. Normalization, therefore, is reckoned in the present research as a potential socio-culturally variable dimension that could be used to measure and even reconstruct the underlying translation norms.

In accordance with its norm-induced nature, the existence of normalization does not necessarily predetermine the translation behavior as absolutely conforming to the target language and culture, and similarly, the existence of non-normalization does not indicate the non-existence of normalization. The difference between normalization and denormalization is only a matter of degree. They constitute two poles of a continuum. Therefore, normalization is used as a relative notion in the present research, and a higher degree of normalization implies a lower degree of “denormalization”.

Based on the above discussions, we could tentatively make the following assumptions:

- (1) There will be observable and measurable evidence of normalization as a surface-structure realization of textual norms in translated texts;
- (2) The degree of normalization in translation will vary with time. Normalization and denormalization might be found to coexist in translated texts as two competing tendencies that pull in different directions, mainly due to the competing power relationship between source and target languages and cultures as well as the tension between faithfulness and idiomaticity that every translator faces.
- (3) The variation of normalization could be contextualized in terms of extratextual norms.

As a surface-structure realization of textual norms, normalization could be generalized based on inferences drawn from observations of large quantities of data, and it is generally accepted that corpus-based studies will reveal “regularities of actual behavior” (Toury 1995: 265), which may be evidence of normalization (i.e., conformance to certain textual norms) or denormalization (i.e., alienating from the norms).

1.3 Overview of Chapters

This book falls into eight chapters. It opens with an introduction to this study, including the research approach and objectives, conceptualization of normalization and structure of the book. The second chapter presents an overview of previous studies on features of translation, and normalization in particular, with a critical review of the corpus methodology in translation studies. Chapter Three is where the empirical research begins.

Construction of a diachronic corpus for the research and the major corpus research tools are described in detail. In Chapters Four, Five and Six we combine a quantitative and qualitative approach to the linguistic manifestations of normalization on different levels, how they change across times and to what extent such changes influence modern Mandarin Chinese. Chapter Four describes the empirical investigation of normalization on the lexical level, including measures like POS distribution, high-frequency words, word-length, and compositionality of quasi-affixes. Chapter Five offers a detailed investigation into the normalization of collocations in English–Chinese translations. Chapter Six addresses the syntactic manifestations of normalization using three measures: hypotacticalization of translated Chinese, use of passive constructions, and sentence/sentence segment length. Chapter Seven addresses the socio-cultural constraints that conditioned the different manifestations of normalization and translation norms in different historical periods. In Chapter Eight, the methods and findings of the entire study are summarized and discussed, as are the ramifications of the present findings for future studies on this topic.

CHAPTER TWO

CORPUS METHODOLOGY IN TRANSLATION STUDIES

2.0 Introduction

This chapter begins with a review of the theoretical background of corpus-based translation studies and traces its development. It looks at the aims, objectives and methods of descriptive translation studies and corpus linguistics, and examines the general applicability of corpus methodology to translation studies. It also discusses a number of aspects of corpus-based translation studies, investigations on translation normalization in particular, that provide an important methodological basis for the present research, followed by some critical views of corpora methodology in translation studies

2.1 Theoretical Background of Corpus-based Translation Studies

2.1.1 The Descriptive Approach to Translation Studies

Since the 1970s, the concept of equivalence associated with the linguistic-oriented “science” of translation has declined, and scholars have become particularly critical of impressionism, anecdotalism and prescriptivism in translation studies, and of approaches that view translations as idealized, speculative entities, rather than observable facts (Toury 1980: 79–81). Descriptive Translation Studies (DTS), developed in the 1970s in the Western world, provides a new approach to translation studies, which not only broadens the horizon of translation studies, but also shifts its focus to target language and target-oriented culture. Through the 1970s and beyond, DTS emphasized description of what translation was and is, distancing itself from previous approaches that were more concerned with prescribing what translation should be.

DTS, first introduced by James Holmes (1972/2000) and then

developed with an emphasis on “function” by Gideon Toury (1980, 1995), was heavily influenced by Even-Zohar’s (1979) polysystem theory which had its origins in comparative literature and Russian Formalism (see Munday 2001: 14, Kenny 2001: 49). The Formalists are mainly concerned with the investigation of the specific properties of literary material that distinguish such material from material of any other kind. This indicates that they were interested both in the formal characteristics of literariness and the differences between literary texts and other orders of writing. The purpose of the literary properties devices is to shape language and to “defamiliarize” “those things that have become habitual or automatic” (Erich 1969: 76). For translation studies, Russian Formalism offers, above all, a way of thinking about the facts of literature, not as single details in themselves but as they exist in relationship to other facts (Jameson 1972). It is the dynamic relationship which counts most in fully describing literature. Both Even-Zohar (1979) and Toury (1980, 1995) argued in terms of systems and active relationships between the various functions of the literary text. This is the basis of their significant contribution to translation studies.

Polysystem theory stresses that translated literature as a system is worthy of study in its own right, since it is a system taking part in the dynamic interaction between systems in the target culture polysystem, and becomes either central or peripheral, primary or secondary, etc., depending on the overall state of the polysystem (Even-Zohar 1978). Even-Zohar’s polysystem theory transformed translation studies into an investigation of the position of translated texts taken as a whole in the historical and textual systems of the target culture. Toury (1980, 1995) took that transformation as his point of departure. Continuing and building on the polysystem work of Even-Zohar, Toury observed that translational phenomena could ultimately be explained by their systemic position and role in the target culture. According to Toury any descriptive research should focus primarily on the target language product because it is the target or recipient culture which “initiates” the decision to translate and the translation process. The goals which the translation is designed to serve are set in and by the receptor systems, and translators operate first and foremost in the interest of the target culture and not that of the source text or source culture. Other recent translation theorists such as Hermans (1985), Gentzler (2004) and Bassnett (1980) have also preferred to focus on the target rather than the source text and on what the target text can tell us either about a given society at a given time or about the system under which a translation is produced.

With a strong emphasis on empirical data and on the social and cultural

context in which the individual translator worked, DTS researchers take as their object of study translated texts as they actually occur from the point of view of the target culture, and seek to describe and explain regularities or repeated patterns observable in translation behavior as evidence of the operation of hypothesized “norms”. Scholars note that translation norms could be observed and reconstructed by studying two major sources: textual sources (the translated texts themselves) and extratextual sources (e.g., semi-theoretical or critical formulations such as prescriptive theories of translation, statements of persons involved in translation, critical appraisals of individual translations or the activity of a translator or “school” of translators, and so forth). Linguistic manifestations of norms of translation have been the focus of much work in DTS, which might be observed using electronic corpora.

Following the interest in the regularities or hypothesized norms of translation, Toury’s (1995) idea that there might be “general laws” of translation began to take hold. One is the “law of growing standardization” which states that “textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favor of more habitual options offered by a target repertoire” (1995: 268). That is, translators will tend to “accommodate” established norms to a greater degree, especially in societies that consider translation “peripheral” (see also Laviosa 1995: 267–268). The other is the “law of interference” requiring that “in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text” (Toury 1995: 275). Principal conditioning factors of the laws are text type, linguistic level, position of the translated literature in literary polysystem, and the relative prestige of the languages and cultures involved in translation.

Toury had called for a more scientific methodology for the descriptive analysis of source and target texts in an attempt to identify probabilistic tendencies or “laws” of translation. Methodology, therefore, is considered an important dimension of DTS. Munday (2001: 112) summarizes the three-phase methodology Toury (1995) suggests for systematic descriptive translation research:

- (1) Situate the text within the target culture system, looking at its significance or acceptability.
- (2) Compare the ST and the TT for shifts, identifying relationships between “coupled pairs” of ST and TT segments, and attempting generalizations about the underlying concept of translation.
- (3) Draw implications for decision-making in future translating.

The descriptive methodology that Toury proposes involves a gradual inductive progression from observable phenomena realized in real

translation products to the non-observable factors that govern translational behavior (Laviosa 2008: 122). Despite criticism of its over-emphasis on the text, as well as its target-orientedness (Hermans 1995), certain scholars of literary translation have found the descriptive approach particularly productive, especially when the corpus invites a longitudinal and historical approach (Merkle 2008: 175). According to Gentzler (2004: 133–134), Toury's theory has had an important impact on translation studies in four aspects: the abandonment of one-to-one notions of correspondence as well as the possibility of literary/linguistic equivalence; the involvement of literary tendencies within the target cultural system in the production of any translated text; the destabilization of the notion of an original message with a fixed identity; and the integration of both the original text and the translated text in the semiotic web of intersecting cultural systems (see Munday 2001: 117).

Toury called for a whole new methodology that would make individual studies transparent and repeatable. The corpus-based methodology clearly links with the descriptive perspective, specifically with the emphasis on the study of the translation product, the interest in identifying typical features of translation, and the combining of quantitative data with close critical analysis of the texts in their socio-cultural environment (Munday 2006). Baker (1993) first saw the potential for the application of corpus methods to the study of translated texts, and since the idea was presented in her early work in the area (Baker 1993, 1995, 1996), Toury's notion of "translation norms", i.e., of socio-cultural constraints regulating the behavior of professional translators and leaving traces in translated texts, along with the more controversial notion of "translation universals" (see section 2.3.1), has been the object of substantial corpus-based studies in particular.

2.1.2 The Corpus Linguistic Approach

Since the 1980s, corpus linguistics (CL) has increasingly gained in popularity. Corpus-based analysis is firmly empirical and typically makes extensive use of computers for analysis of the actual patterns of use in a large and principled collection of natural texts (see Biber, Conrad and Reppen 1998: 4).

Corpus research has become a key element of almost all language studies. The focus of corpus linguistics is on the meaning which is being verbally communicated between the members of a discourse community. Frequency is an important parameter for making general claims about the discourse. According to Stubbs (2001: 151), corpus linguistics is concerned

with “what frequently and typically occurs”, as opposed to isolated, unique instances of language. It investigates relations between frequency and typicality, and instance and norm. Similarly, Kennedy (2000: 7–10) observes that corpus research enables generalizations to be made about language use, and that the interest is not just in what occurs but in what is probable—“what is likely to occur in language use”. It is thus particularly helpful in providing “big picture” perspectives on discourse—determining patterns of language behavior across many texts, identifying typical and unusual choices by users, and describing the interactions among multiple variables (Conrad 2002).

Corpus linguistics looks at language from a social perspective, which necessarily requires a combination of both quantitative and qualitative analytical techniques for descriptive language study, though research using corpora is generally seen as empirical, data-based or data-driven. Advances in computer technology have made it easier to work with larger quantities of text; however, they have not drastically changed the nature of text-based linguistic study: “corpus linguistics is not a mindless process of automatic language description” (Biber *et al.* 1998: 2), and “the goal of corpus-based investigations is not simply to report quantitative findings” (Biber *et al.* 1998: 5). Rather, a crucial part of the corpus-based approach is going beyond the quantitative patterns to propose qualitative, functional interpretations of the linguistic features (Biber *et al.* 1998). As a result, a large amount of effort in corpus-based studies is devoted to explaining and exemplifying quantitative patterns.

Corpus linguists thus take an approach to linguistic study that is consistent with the empiricism advocated in descriptive translation studies, in contrast to the use of introspection based on intuitive data or a priori assumptions. With its emphasis on typicality in actual texts and a combination of quantitative and qualitative approaches, corpus linguistics clearly links with the methodology of DTS.

Kenny (2001) summarizes the common ground corpus linguists share with scholars working in the area of descriptive translation studies: both groups insist on the primacy of authentic data—actually occurring text in corpus linguistics, and actually occurring translated text in descriptive translation studies—and both are concerned to identify recurring patterns in texts, and to relate these patterns and deviations from them, to features of the wider context of situation. Given the common ground they share, it is not surprising that “the two are beginning to converge in what has been dubbed corpus-based translation studies” (CTS) (Kenny 2001: 48).

The discipline of CTS now offers researchers the opportunity for first-order observation of any “regularities” or “recurrent features” produced

by the act of translation. Corpus linguistics clearly has certain analytical strengths within the broad theoretical framework of DTS, since it provides a method for the description of language use in translation, whether this concerns the target text only or both source and target texts in parallel. Emphasis on the descriptive (especially in opposition to the prescriptive) is undisputed (Olohan 2004: 16).

2.2 CTS: Methodology and Development

CTS, growing out of corpus linguistics and thus inherently belonging to linguistic approaches to translation, is an area of research that is growing in prominence in translation studies. Since the 1990s, the rapid development of corpus linguistics in general, and the development of multilingual corpora in particular, have brought even more vigor into Descriptive Translation Studies (Xiao *et al.* 2010), and there has been a growing interest in the application of computer-assisted methods of investigation to the study of translation and translated texts. The idea of using corpora in translation studies research was first put forward by Baker (1993), and has since been developed in her own research (e.g., Baker 1995, 1996) and in Laviosa's (e.g., 1998a, 1998c, 2001). The availability of large corpora of both original and translated text, together with the development of a corpus-driven methodology, will enable translation scholars to uncover the nature of translated text as a mediated communicative event (Baker 1993: 243). Following Baker, a large body of research in translation has adopted a corpus-based methodology to try to shed light on the "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker 1993: 243).

Corpus-based studies have been conducted mainly using two kinds of resources: the more traditional parallel corpora, made of source texts in language A and their translations into language B, and monolingual comparable corpora, made of originals in language A and comparable translations into language A. Parallel corpus approaches are more appropriate for the analysis of local shifts and strategies (Bernardini 2007). Studies following this approach have focused, e.g., on explicating shifts (Øverås 1998), on normalizing/sanitizing shifts (Kenny 2001) and on translator choices with implications for a description of translator's style (Malmkjær 2004).

There has been more stress on the use of the monolingual comparable corpus since Baker suggested a "shift in the focus of theoretical research in the discipline, a shift away from comparing either ST with TT or language

A with language B to comparing text production per se with translation” (Baker 1995: 233). The monolingual comparable corpus herein “consists of a corpus of translations and comparable non-translations in the same language” (Olohan 2004: 35). The authentic, non-translated texts are chosen from the target language repertoire and represent the genre to which the target text is supposed to belong. The first corpus of this kind, designed and constructed under the direction of Baker, was the *Translational English Corpus*), which consists of a collection of written texts translated into English from a range of different source languages, European and non-European, with an overall size of currently around 10 million words. TEC consists of four subcorpora: fiction, biography, news and inflight magazines, usually used in comparison with a comparable set of non-translational texts taken from the *British National Corpus* (BNC).

Comparable corpora have been used to investigate the various translation phenomena, and attempt to give answers to questions such as: Are there translation norms or translation universals? (Baker 1993: 243–245) Following Baker, comparable corpora has been one of the fundamental concepts in corpus-based translation studies and is believed to be “promising resources for the study of collocational restrictions in translated vs. non-translated language” (Bernardini 2003: 90). Monolingual comparable corpora have been used extensively to compare overall textual features such as sentence length, lexical variety, ratio of content words to function words (Laviosa 2002), or more specific patterns of use of (semi-)grammatical (Olohan 2001, Olohan and Baker 2000) and lexical features (Tirkkonen-Condit 2004, Mauranen 2000). The aim of most research using a comparable corpus of this kind is to capture “patterns which are either restricted to translated text or which occur with a significantly higher or lower frequency in translated text” (Baker 1995: 235), which can help find out about “the nature of translated text in general and the nature of the process of translation itself” (Baker 1995: 236).

Translation, as a complicated communicative activity, involves considerations of culture, context and socio-political factors. Naturally, corpus-based study of translation, as with corpus linguistics, requires a combination of both quantitative and qualitative methodology to explore the pragmatic factors related to the production of translations. As Baker (1998) argues, the aim of developing a corpus-based methodology “is not merely to unveil the nature of the ‘third code’ per se, but most importantly, to understand the specific constraints, pressures, and motivations that influence the act of translating and underlie its unique language” (cited from Laviosa 1998b: 475). Sinclair (2005) also cautions against the danger

of a vicious circle of researchers constructing a corpus to reflect what they already know or can guess about its linguistic detail. Li and Zhang (2010) observe that a useful and sensible corpus-assisted translation research must have a phase of sense-making following the presentation of statistical results, as shown in Figure 2-1:

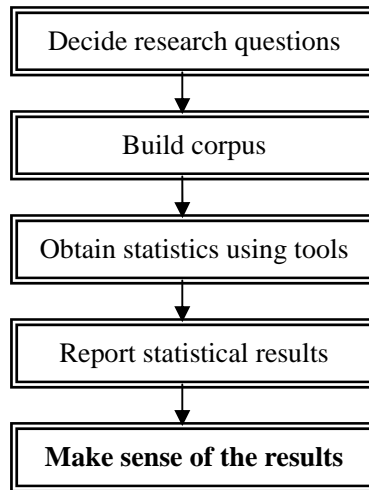


Figure 2-1 Flowchart of a corpus-assisted translation study (Li and Zhang 2010: 246)

According to Li and Zhang (2010: 247), simply providing the numbers does not tell us much about the process of translation. What is really important and useful is the part of sense-making, essential to any corpus-assisted translation study. The step of sense-making is to go further to explain the statistical results by looking at the causes for such tendencies as revealed in the statistical results, and answer questions such as how and why the translation came about the way it did, and what social, cultural and political effects it brought about in the target language.

It can thus be seen that CTS builds upon the studies of scholars working in the descriptive approach to translation studies, and those of scholars working with corpora themselves. By describing other texts and language production in terms of their relationship or similarity to translation, the traditional “translation-as-derivative” bias is reversed (Olohan 2004: 36). With the view of translation not as an individual act of transfer of a source text into a target language, but rather as a socio-culturally regulated communicative event in the target language

community, CTS “marks a turn away from prescriptive approaches to translation toward descriptive approaches, approaches developed by scholars, notably by polysystems theorists such as Itamar Even-Zohar, Gideon Toury, and André Lefevere” (Tymoczko 1998).

2.3 Aspects of Corpus Research on Translational Language

Up to the present, corpus-based translation studies has primarily been concerned with describing translation both as a process as well as a product, mainly by comparing corpora of translated and non-translational native texts in the target language, especially translated and native English. Some studies focused on features of translational language, while others examined translator’s styles.

2.3.1 Features of Translational Language

Since Baker (1993) promoted the use of corpus tools for exploring the linguistic features that render the language of translation different from the language of non-translated texts, the search for general laws and regularities has received a new impulse. Baker (1996) further posits a number of features as “universal features of translation”—“the constraints which are inherent in the translation process itself” (Baker 1993: 242)—most noticeably simplification, explicitation, and normalization. Since then, the majority of product-oriented translation studies have attempted to provide evidence to either support or reject the so-called “Translation Universals” (TU) hypotheses that are concerned with features of translational language as the “third code” (Frawley 1984), which are supposed to be different from both source and target languages and independent of any particular language pair, text type, translator, or historical period. Following Baker, researchers such as Øverås (1998), Laviosa (1998a), Olohan and Baker (2000), Puurtinen (2003b), Tirkkonen-Condit (2004), among others, have made further efforts in developing ways to test as well to propose new hypotheses about translation universals, and relevant differences have been found between the lexical and syntactical make-up of translated and non-translated texts across various languages.

Explicitation relates to the tendency in translations to “spell things out rather than leave them implicit” (Baker 1996: 180). Øverås (1998) investigates the explicitation hypothesis using the English–Norwegian