Exploratory Statistical Techniques

for the

Study of Literary Translation

by

Meng Ji

Table of Contents

Tables and Figures

Figures

Appendixes

Introduction

This book is divided into four chapters which present three illustrative corpus studies of Chinese translations. A shared methodological feature of the case studies is the use of exploratory statistical techniques in the quantitative processing of translations. The three case studies are selected deliberately. First of all, they touch upon three different language combinations which are (1) Colombian Spanish and Chinese; (2) English and Chinese; (3) Japanese and Chinese. Secondly, the case studies presented cover several textual genres including mystery and detective fictions, general fictions and religious texts. The three case studies prove that exploratory analytical techniques including principal component analysis and hierarchical clustering analysis may be widely applied in empirical translation studies. The statistical results produced shed new light on Chinese translations of diverse textual genres involving different language combinations.

Chapter 1 focuses on the study of stylistic variation in translation from a cultural and genre perspective. It argues that translation of culture-specific textual genres into a distinct language and cultural system often entails important linguistic experiments. This is shown in important stylistic variations between two first early Chinese translations of Garcia Marquez's *One Hundred Years of Solitude* by Gao Changrong and Huang Jinyan in 1984. The language and cultural experiments are reflected in the shifting genre of the Chinese translations from the Spanish original.

The corpus analysis suggests that on the one hand, Gao's first indirect translation of *One Hundred Years of Solitude* via English shows more similarities to Chinese translations of detective and mystery fictions; on the other hand, Huang's first direct translation from Spanish is more related to Chinese translations of romance fictions. The shifting genre of the Chinese translation of *One Hundred Years of Solitude* is defined as genre-shifting in this book, which represents a unique and latent translational phenomenon that has been rarely discussed in translation studies.

Genre shifting is particularly relevant to experimental translations. The discovery of this translation phenomenon has important implications for empirical translation studies, especially with regards to the use of comparable and/or reference corpora. In corpus translation studies, the use of reference corpora is normally based on the assumption that the textual genre of the translation is consistent with that of the source text. The identification of genre shifting challenges this view by questioning its validity with experimental translations like the two case studies presented in Chapter 1 and Chapter 2.

Genre shifting in literary translations requires a priori test of the strength

of relationship between an experimental translation and existing literary genres. This can be an extremely time-consuming process that is error-prone, given the extensive cross-corpus statistical analysis entailed. To solve the technical difficulties presented by genre shifting, Chapter 1 introduces two typical exploratory statistical techniques, i.e. principal component analysis and hierarchical clustering analysis in the study of stylistic variations in literary translations. The methodological significance of exploratory techniques is that they are essentially exploratory in the processing and modelling of large-scale translational corpus data. In other words, they can effectively detect the strength of correlation between different corpora.

In Chapter 1, the use of hierarchical clustering analysis found that Gao's first indirect translation had the smallest distance score from Chinese translation of detective and mystery fictions. A small distanc score indicates a high level of similarity between Gao's translation and Chinese detective translations. By contrast, Huang's version of the Colombian novel had the smallest distance score from Chinese translations of romance fictions. The corpus statistics pointed to the shifting of genres between the first two Chinese translations of Garcia Marquez's Nobel Prize winning novel of magical realism.

Chapter 1 introduces principal component analysis to study the distribution of a range of part-of-speech taggers in nineteen corpora of translational and original Chinese including Gao's and Huang's translations of *One Hundred Years of Solitude* published in 1984. It was found that at least four part-of-speech categories were recurrent and defining features of Huang's translation. They were Chinese lexical pattern ZHI (之); semicolon; monosyllabic morphemes as Chinese numerals and the lexical pattern Dehua (的话). Based on the corpus finding of the similarity between Gao's translation and Chinese translations of detective fictions, a list of key part- of- speech categories in Chinese detective translations was extracted by using relevant corpus mining software.

The keyword list compilation drew on the comparison between the Zhejiang University Corpus of Translation Chinese and its sub-corpus of Chinese detective translations. The statistical technique used was log-likelihood test as an inbuilt function of the CQP multiple corpora search platform. These part-of-speech items of high-frequency in Chinese dectective translations were then explained and analysed in light of translational universals, genre-specific translation features, source text impact and the translator's stylistic use of language. Textual excerpts were taken from Gao's translation as illustrations of textual features typical of Chinese detective and mystery translations.

Chapter 1 adopts an essentially corpus-driven approach to the study of stylistic variations in translations. It challenges the hypothesised correspondence between the text genre of the source text and that of its translation into the target language. The use of exploratory statistics played an instrumental role in the

detection and modelling the relationship between translational corpora. In the case study, the use of exploratory techniques effectively identified the similarities and dissimilarities between the Chinese translations of *One Hundred Years of Solitude* and the Chinese reference corpora used. In doing so, Chapter 1 advanced the corpus-driven approach to the study of translation stylistics.

Chapter 2 tests the validity and productivity of exploratory statistics in the study of the translation of conventional text genres. It applies exploratory statistical techniques in the study of the translation of a Japanese detective novel, *Confessions* by Minato Kanae (2010) into Chinese. The corpus analysis found that genre shifing widely exists in translations, which include translations of experimental fictional genres such as magical realism; and translations of conventional text genres such as detective and mysterious fictions. Genre shifting in translation was seen as the result of the use of unusual narrative techniques in the source text.

Chapter 3 moves away from stylistic variations in translations to the detection of consistent and underlying patterns in translations, which are known as translation universal features in descriptive translation studies. Chapters 1 and 2 argue that the lack of advanced research methods has hampered the study of stylistic variations in translations. In a similar fashion, the development of theoretical hypotheses such as translation universals requires the introduction of advanced statistical methods from related research fields like applied linguistics to empirical translation studies. The use of exploratory techniques aids in the verification of existing hypothetical claims such as translation universals; or the discovery of new textual patterns.

The main purpose of Chapter 3 is to introduce and test exploratory statistical techniques including principal component analysis in the study of translational universals. Chapter 3 studies the systematic variations between two sets of corpora of translational and original Chinese. The comparison of the corpora was conducted at a cross-genre level. Instead of approaching translation universeals from an inter-lingual perspective, Chapter 3 examines translation universal features in a variety of text genres of translational Chinese. The existence of textual and linguistic features that may be related to translation universals is tested at a cross-genre level.

This involves the comparison of a range of genre-specific sub-corpora of the original Chinese corpus LCMC and its counterpart, the ZJU Corpus of Translational Chinese. In Chapter 3, the use of exploratory techniques identified several important textual patterns. First of all, principal component analysis verified the existence of translational universals such as explicitation and enhanced textual cohesion across a range of Chinese translational texts, especially in Chinese translations of academic prose, official documents and reports and general fictions.

Secondly, the deployment of exploratory techniques identified a new feature underlying Chinese translations. In Chapter 3, this new textual feature was defined as enhanced idiomaticity, which is existent at both the semantic and the syntactic levels. This is reflected in the recurrence of linguistic and textual devices in a variety of Chinese translations. At the semantic level, enhanced idiomaticity includes the increased use of Chinese idiomatic and fixed expressions; and multi-character words in Chinese translations. At the syntactical level, enhanced idiomaticity in Chinese translations is shown in the increased use of grammatical devices and verbal structures typical of Chinese syntax, especially those borrowed from historical Chinese. These include Chinese quantifiers and typical Chinese verbal structures including 所 (suo) ,把 (ba) ,被 (bei) ,是 (shi) and 有 (you).

Chapter 3 illustrates that exploratory techniques and multi- dimensional analyses can be used to verify the existence of translational features such as explicitation and enhanced textual cohesion. Morever, the use of advanced textual modelling techniques is instrumental in the identification of new textual features that underlie a variety of translations. The new translation phenomenon detected in Chinese translational corpora is defined as enhanced idiomaticity. This new textual feature proved to be consistent in a range of Chinese translation genres; and it holds important implications for the teaching and research of translations.

In conclusion, I argue that to advance empirical translation studies, the introduction of and experimentation with exploratory statistical techniques such as principal component analysis and clustering techniques is essential. The case studies presented show that the deployment of advanced statistical methods may not only help verify existing theoretical hypotheses such as translation universals and norms; but more importantly, they may assist us in the detection of new translational phenomena such as genre shifting and enhanced idiomaticity, which in turn helps to broaden the scope of current empirical translation research.

The systematic application of exploratory techniques, as the case studies illustrated, advances current corpus translation studies. In many previous and current translation studies, the use of translational corpora and corpus metods is mainly for the purpose of verifying and testing presumed theoretical hypotheses. By contrast, with the introduction of exploratory techniques, the focus of corpus analysis is the modelling and extraction of recurrent textual patterns in translational corpora. The regular patterns or consistent features revealed in translations provide a solid empirical basis for the formulation of new theoretical frameworks and models for the rapidly growing field of corpus translation studies.