Quantitative Exploration of Historical Translations

A Corpus Study of Tetsugaku Jii

by

Meng JI

RAM-Verlag

2013

ISBN: 978-3-942303-16-3
© Copyright 2013 by RAM-Verlag, D-58515 Lüdenscheid

RAM-Verlag
Stüttinghauser Ringstr. 44
D - 58515 Lüdenscheid
RAM-Verlag@t-online.de
http://ram-verlag.eu

Table of Contents

Introduction 1

Chapter 1
Corpus Study of Historical Translation 4
Chapter 2
Research Database of Historical Translations 9
Chapter 3
Linear Regression Analysis of the Use of Suffixes in Tetsugaku Jii 12
Chapter 4
Agglomerative Hierarchical Clustering of Length-Specific Tokens in Tetsugaku Jii 19
Chapter 5
Principal Component Analysis of Length-Specific Tokens 22
Chapter 6
Conclusion 29

References 31

Appendix 1
Annotated single character expressions in Tetsugaku Jii (Inoue 1884) 34
Appendix 2
Annotated two-character expressions in Tetsugaku Jii (Inoue 1884) 36
Appendix 3
Annotated three-character expressions in Tetsugaku Jii (Inoue 1884) 49
Appendix 4
Annotated four-character expressions in Tetsugaku Jii (1884) 64
Appendix 5
Annotated five-character expressions in Tetsugaku Jii (1884) 80
Appendix 6
Distributional statistics of annotated corpus data 85
Appendix 7
Selected scientific suffixes in characters in Tetsugaku Jii (1884/1912) 87

Tables and Figures

Table 1 Distribution of length-specific tokens in *Tetsugaku Jii* 14
Table 2 Skewness check of the original data set 16
Table 3 Normality check for English suffixes 17
Table 4 Normality check for Chinese suffixes 17
Table 5 Linear regression analysis of the use of suffixes in *Tetsugaku Jii* 18
Table 6 Proximity matrix of Agglomerative Hierarchical Clustering 20
Table 7 Principal component analysis result 24
Table 8 Factor loadings of observational variables after Varimax rotation 24
Table 9 Dimension Explanation 28

Figure 1 Dendrogram of AHC
(on length-specific tokens in Inoue's translation) 21
Figure 2 PCA Scree Plot 23
Figure 3 Principal component analysis of observational variables 26
Figure 4 Observations (cases) after Varimax rotation 27

Introduction

The main purpose of this book is to test the validity and productivity of advanced corpus methodologies in the study of historical translations, particularly historical dictionaries. It focuses on the translation of Western scientific works in Japan and China in the early modern period, particularly in the nineteenth century. The corpus analysis presented is highly interdisciplinary which draws upon research methods adapted from corpus linguistics, contrastive linguistics (English/Chinese/Japanese), comparative cultural studies and so on. The detailed corpus analysis aims to bring valuable insights into the dynamics and shifting patterns of the cross-cultural and cross-linguistic interactions between Japan, China and the West in the early modern period.

The key question that this book aims to address is how early modern Western scientific idea sets and concepts were introduced, translated and assimilated into the native language and cultural systems of Japan and China during the early modern period. Modernization in East Asia during the early modern period was an extremely complex process that can be hardly explained by a simplicist borrow-to-use model which was influential in the study of early cultural contacts between the West and East Asia. In this book, I argue that the modernization of East Asia entails the experimentation with Western scientific and cultural imports to meet the needs and demands of the social transformation and development of East Asian cultures and societies.

This is centrally reflected in the development of working translation models in influential early modern Japanese and Chinese translations. As this book will show, the development of working and viable language models for the translation of Western scientific input underscores the inherent systematic differences between Japanese, Chinese and Western languages. The distinct translation tactics used by different translators largely reflect their different and sometimes confrontational ideological stances and strategic views over the transformation and modernization of East Asian languages in the early modern period.

In this book, the study of historical scientific translations focuses on a highly influential early English Japanese dictionary *Tetsugaku Jii*, also known as *A Dictionary of Philosophy*. It was first published by the renowned Japanese linguist and philosopher Inoue Tetsujiro (1855-1944) in 1881 and reprinted in 1884. The third edition of the dictionary was printed in 1912.Through the exploration, processing and analysis of purposely collected historical translation data, this study aims to identify textual and contextual factors which explain the development of new scientific expressions in *Tetsugaku Jii*, which laid the foundation of the early modern Japanese and Chinese scientific terminologies. A large number of scientific terms and expressions created in *Tetsugaku Jii* remain current in modern Japanese and Chinese scientific writings (Zuo, 2000).

As with any corpus-driven empirical translation study, I first extracted a large number of translated terms and expressions in the English and Japanese bilingual dictionary *Tetsugaku Jii*. Next, I proceeded to compare and analyse different translated terms and expressions by focusing on their lexical and syntactical features, for example, the length of translated scientific words in

characters, and the use of scientific suffixes in newly created scientific expressions. The corpus-driven analysis identified a variety of source textual features that largely explain the creation and development of length-specific scientific expressions in *Tetsugaku Jii*.

The corpus approach to the study of historical translations helps to systematise the processing and analysis of a large-amount of historical textual data. The use of advanced statistical methods can not only identify useful textual and linguistic patterns in historical translations, it can also formulate and verify theoretical hypotheses regarding the production and reception of translations in the target language and cultural background. In doing so, this book aims to advance the empirical and corpus-driven study of historical translations, especially historical bilingual dictionaries.

The methodological focus of this book is the application of and experimentation with exploratory statistical techniques in the study of historical translation materials. A distinctive feature of exploratory techniques is that it can identify the relationship between the source and the target text; and verify the strength of the relationship between the text pair. To be specific, in the current study, the use of relevant statistical methods, multivariate analysis in particular, explores a range of textual and linguistic features of the English source language, with a view to explaining the distribution of length-specific tokens in the first edition of *Tetsugaku Jii* published in 1884.

This book introduces three statistical techniques in the study of *Tetsugaku Jii* which are linear regression analysis, hierarchical clustering techniques and principal component analysis. The use of these three different techniques helps to answer three different types of research questions which are first of all, whether there is any significant dependence relationship between the source text and the target text expressions (linear regression analysis); secondly, how the different types of character expressions in *Tetsugaku Jii* differ from each other; and lastly, what are the source text features that motivated the development of new scientific expressions of varying character lengths in the Japanese translation (principal component analysis).

Linear regression analysis of the corpus data showed that the consistent use of mono- syllabic and disyllabic suffixes in the Japanese translation might be reasonably explained by relevant source text language features, for example, the use of suffixes of Latin or Greek origins in English scientific expressions. Hierarchical clustering technique found that the use of two- and three-character expressions and words in *Tetsugaku Jii* were very similar to each other. Newly created two- and three-character words represent the majority of new scientific terms created in *Tetsugaku Jii.*

The proliferation of two- and three-character words in early modern Chinese is closely related to the lexical evolution of historical Chinese. Ji (2013) finds that while the use of mono-syllabic or single-character words is predominant in historical Chinese, the percentage of two-character and three-character expressions picks up in modern Chinese, especially since the nineteenth century. It was suspected that the translation of a large number of Western scientific works in the nineteenth century underpins lexical changes and variations in early modern Chinese.

This hypothesis was proved in the statistical analysis. Hierarchical clustering analysis detected a higher level of similarity between single-character words and two- and three-character words in *Tetsugaku Jii*, which had a huge impact on the development of modern Chinese lexis. Two- and multi-character expressions tend to be more explicit and descriptive than single-character words. Lastly, principal component analysis of the corpus data showed that the creation and development of length-specific tokens in *Tetsugaku Jii* was associated with specific source text features such as the subject fields of source text expressions, their part-of-speech functions and the length of English source text expressions in terms of the constituent words.

This book demonstrates that the use of exploratory statistical techniques may not only identify novel and revealing textual and linguistic patterns in translational corpora; more importantly, such versatile statistical techniques can be fully exploited to develop new lines of empirical inquiries which will point to new directions for historical translation research.