Most advocates of descriptive empirical research would agree that knowledge comes from empirical analysis. For empiricists, knowledge or ideas that cannot be proven in practice are only speculative and their work is mainly driven by their desire to turn theoretical speculation into empirical and measurable facts. In the field of translation, corpus-based translation studies have become one of the best representatives of this type of descriptive empirical research, adopting a more experimental and rigorous methodology that can provide data suitable to be analysed with statistical tools. And the task is by no means straightforward. As in linguistics, translation researchers work with concepts, such as language or culture, which are inherently fuzzy and difficult to apprehend and reduce to simple variables. Besides, the translation act involves a re-codification process into a different language that increases the complexity of the task and makes it more difficult to formulate research questions in terms of variables that are susceptible of objective and quantitative study. But despite difficulties, corpus-based translation studies have made successful attempts at moving towards more systematic and quantitative methods of analysis (cf. Baker 1996; Laviosa 2002; Olohan 2004; Beeby et al. 2009; Zannetin 2012).

The book reviewed here illustrates precisely the move of translation research in the direction of corpora studies by providing – as the editors themselves...
outlined in the preface – ‘a comprehensive guidebook to the essential quantitative methods in corpus-based translation studies (CBTS)’ (p. vii). The foremost motivation for the book springs from a methodological limitation that still constitutes a major hindrance in the advancement of quantitative translation research, namely, the scarce statistical background of scholars mostly trained in linguistic and translation competences. As a consequence, most of the studies carried out in CBTS have been essentially exploratory and the area is still in need of developing adequate quantitative methods of analysis that allow researchers to test theoretical models in the practice. This book is an attempt to palliate this situation by introducing a number of case studies that illustrate the application to translation research of some of the statistical tests used in corpus linguistics.

The papers in the volume are classified into four different sections, which mark the progression from theoretical aspects to empirical work. Thus, while the first two sections still focus on some theory-driven aspects of research design, the last two dive into practical work delving into the exploration of some stylistic and lexical features of literary translations. Truth be told, even the most theoretical sections are empirically biased in a worthwhile effort to close the gap between theory and practice.

The first section comprises three papers devoted to exploring the interplay between theoretical aspects and empirical findings. The article by Barbara Lewandowska-Tomaszczyk provides a method to measure translational equivalence by establishing a text profile of original text and translation based on qualitative (e.g. perceptual, functional, emotional factors) and quantitative criteria (e.g. keyness and collocation patterns). The fact that equivalence is no longer conceived of as a static relationship but as a set of re-conceptualizations of the original scene construal is one of the most valuable contributions of the method proposed. The article by Stefan Gries and Stephanie Wulff describes how theoretical translation models can be tested using the statistical method of regression analysis. For readers who are not familiar with this method, their paper provides a rather technical – but still graspable – and in-depth account of how to compute simple binary logistic regressions and linear regressions with the open source programming language R. One of the main fortes of the paper is the fact that the different types of regressions explained are exemplified by using data taken from the INTERSECT corpus of English texts and their translations into French and German. The third article of this section, the paper by Meng Ji, focuses on hypothesis testing by comparing two different translations of Cervantes’ *Don Quijote* into Chinese. This paper proposes a quantitative approach to explore the various relationships that underlie a translation, namely, the relationship between source text (ST) and target text (TT), between TT and target language (TL) and culture (TC) and between
the style of TT and the diachronic development of TL in general terms. As in the two previous studies, the application of the method is illustrated with an empirical investigation, which explores the relationship between ST and TT by mapping out the distribution of archaisms in the three texts under study. In the second part of the study, a similar quantitative methodology is used to establish the relationship between the change in translated language and the diachronic development of modern Chinese by mapping out the distribution of idioms in one of the translations analysed and two large-scale modern Chinese corpora.

The second section of the book contains three other papers that introduce essential aspects of corpus statistics, such as the construction of a bilingual (Spanish and Norwegian) translation corpus (as in the article by Lidun Hareide and Knut Hofland), the different ways of describing a text by quantifying the frequencies of linguistic features (as in Michael Oakes’ paper) and the various clustering techniques that can be used to group translated texts and interpret information about the similarities detected (as in the paper by George Ke). Despite the authors’ efforts to use examples that illustrate theoretical concepts, the weight on theory is greater than on practice, and the contents discussed in this section focus, to a higher degree, on design fundamentals and technicalities of the discipline. Some of the aspects introduced here may in fact appear rather arid to readers who are not familiar with statistical methods. But we agree with the authors that they are essential to understanding the practicalities of corpus statistics and therefore deserve a place in the spotlight.

The third part of the book introduces three studies that illustrate the use of various statistical techniques to explore three key questions in literary translation: the possibility to discriminate between different translations of the same original text, the degree of consistency between original text and translation and the possibility to measure individual translators’ styles. The paper by Michael Oakes and Meng Ji demonstrates how bivariate statistical tests can be used to discriminate between different translations of the same ST. Much of the practical value of their work lies in the wide range of statistical tests applied (seven different tests) and in their application to analyse a suitable array of features: they use the Student $t$ test for matched pairs to compare translations in terms of sentence length and emotion words, the Student $t$ test for independent samples to compare the rate of occurrence of emotion words and the chi-squared test to compare multiword phrases. They also apply Pearson’s $r$, Spearman’s Rank Correlation Coefficient and Wilcoxon’s signed ranks matched pairs test to differentiate between translations in terms of sentence length, and the Mann Whitney $U$ test to compare translations in terms of type-token ratio. The work by Jon Patton and Fazli Can contributes to explore the degree to which a translation deviates from the original text by applying
stylometric analysis. They demonstrate how this technique serves to establish the degree of consistency between original text and translation by comparing five style markers: sentence length, most frequent words, word length in text, word length in vocabulary and type-token ratio. The section closes with the article by Jan Rybicki, which addresses another major concern of translator scholars, the question of the translator’s invisibility from a quantitative point of view. Rybicki shows how the stylometric distance method known as Burrow’s Delta – generally used for discriminating authorial ‘uniqueness’ – proves inadequate for differentiating between individual translators’ styles. The analysis carried out reveals a consistent tendency of the translations under study to cluster by author and volume rather than by translator. This result has interesting implications for translation studies, since it indicates that two translations of the same text into the same language share much more than any other two literary texts written in the same language.

The final section of the book comprises four papers that shift the focus of study from literary translation to the analysis of translation lexis in general. This part also addresses some relevant issues for translation studies from a lexical and terminological point of view, such as the translation of scientific terminology, ambiguous terms or derivational affixes. The section opens with the study by Meng Ji, which tests the validity of statistical procedures like hierarchical cluster analysis for the study of historical scientific translations. The analysis identifies relevant patterns in the development of key linguistic features (such as token length and functional particles) in modern Chinese scientific language. The paper by Alexandre Sotov applies a branch of mathematics known as game theory to analyse the relationship between translators’ use of techniques, such as transcription and explicitation, to render ambiguous Vedic terms and source text content. He proposes a series of models of the translation process that serve to illustrate the ‘adaptive’ nature of translators’ behaviour, showing how translators learn to adapt to the source text environment through a large number of trials. The work by Gard Jenset and Barbara McGilivray applies three multivariate techniques (i.e. factor analysis, principal component analysis and correspondence analysis) to investigate whether source language and translators’ background can influence affix productivity and use. The analysis reveals that translators’ choice in affix use is more constrained by text type conventions than by source language or translators’ background. The section closes with a paper that explores one of the central issues of modern translation studies, namely, the degree to which translated texts differ from non-translated texts in the same language. The paper by Gert De Sutter, Isabelle Delaere and Koen Plevoets combines profile-based correspondence analysis and logistic regression modelling in order to determine the impact that different Dutch varieties or lects have on the lexical choices made
in translated versus non-translated texts. Their hypothesis predicts that translated texts contain more formal lexemes than non-translated texts. As previous work on translation universals has suggested, their results indicate that translated texts display features that differ significantly from those of non-translated texts. Nevertheless, and contrary to what was expected, the differences are dependent on text type and source language.

On the whole, the book constitutes a worthy contribution that underlies the benefits that quantitative corpus linguistic methods can yield to the field of translation studies. Although from the point of view of translation scholars the statistical contents described may seem at times rather technical, the authors manage to keep the interest by suggesting rewarding applications to the analysis of translated language. Moreover, their endeavour to propose feasible empirical studies provides valuable information about the type of linguistic features that are most adequate to explore the different research questions formulated. In this sense, the volume also succeeds in addressing a set of questions that is representative of some of the pivotal issues in translation studies, from the controversial notion of equivalence between original and translation to the role played by individual translator’s styles, different source languages, text type conventions or translation-specific features.

One of the reasons why translation scholars turned their backs on linguistics during the so-called ‘cultural turn’ of translation studies in the late 1980s was certainly the speculative nature of most linguistic approaches, which lacked empirical studies based on real translation data (cf. Snell-Hornby 1988, 2006). However, nowadays the empirical and quantitative base of corpus linguistics can finally pay off the experiential deficit of traditional linguistic approaches. The papers reviewed here give a taste of what might be possible to explore and achieve by applying corpus linguistic methodologies and statistical methods to research questions in translation studies. Perhaps, as De Sutter, Delaere and Plevoets claim in their paper, the time has come for ‘translation scholars to look over the wall and interact with other corpus linguists more frequently’ (p. 343). Or even better, translation scholars and corpus linguists can at last work together to knock down the imaginary wall that has long kept them apart and take a new turn towards a more rigorous and systematic CBTS.

References


**Book reviewed**