

An Empirical Approach to Translation Studies

Amidst the growing body of empirical translation studies and corpus translation studies in particular (CTS), the current volume represents the latest research in key areas of CTS such as machine translation (Chapter 1, Michael Oakes), translation genre variation and shifting (Chapter 2, Meng Ji), translation stylistics (Chapter 3, Defeng Li) and translation universals, including testing of the Gravitational Pull Hypothesis (Chapters 4–5, Lidun Hareide). The structural organization of the book is balanced between theoretical discussion and illustrative case studies. It aims to provide a focused introduction to the research paradigms which prevail in current CTS, i.e., from the development of statistical machine translation to the exploration of recurrent translational patterns called translation universals. From Chapter 1 to Chapter 5, the levels of theoretical postulation increase, as the research methods used gradually move from essentially corpus-driven (Chapter 1 and 2), via corpus-assisted (Chapter 3) to typical corpus-based translation studies (Chapter 4 and 5).

The distinction between these three main research paradigms within the current CTS, which is evolving rapidly, is largely based on the purposes and aims of the use of empirical evidence in the study of corpora. Throughout the book, the frequency-based analysis of language corpora, monolingual or multilingual, plays an instrumental role in the corpus analysis of translation. In corpus-driven translation research as exemplified by Chapter 1 (on statistical machine translation), and from a different perspective by Chapter 2 (on genre studies), corpus analysis tends to focus on the statistical modeling of linguistic and textual patterns which lead to the development of new computational language models, conceptual dimensions and analytical instruments in translation studies.

Chapter 1 offers an overview of important research paradigms in machine translation, i.e., rule-based machine translation, example-based machine translation, translation memories and statistical machine translation. The significance of this chapter is that it uses case studies in multiple languages to illustrate the rationale behind competing language and translation models. The linguistic analysis is enhanced with detailed explanations of relevant statistical procedures which allow readers to obtain an in-depth understanding of machine translation systems from Google Translate to popular computer-assisted translation (CAT) language resources like translation memories.

Chapter 2 presents a quantitative analysis of contrastive distributional patterns of part-of-speech categories in monolingual English and Chinese corpora, and corpora which contain Chinese translations of English source texts. The corpus study adopts an essentially corpus-driven approach to the analysis of the quantitative data extracted from large-scale language corpora. The statistical analysis constructs three distinctive genre classification models for English, Chinese and translational Chinese as represented by the three large-scale corpora under study. The analysis shows that English written genres have a clear focus on techniques involved in the delivery of textual information. By contrast, the genre system of original

Chinese gives more emphasis to language style rather than the delivery of actual textual information. The focus on the quality and stylistic features of the language implies that the prioritization of the aesthetic value of writing exists widely in the modern Chinese genre system, which is a long-standing tradition in the Chinese language and cultural system.

The exploratory statistical analysis of translational Chinese genres reveals that the genre system of translational Chinese is more complex than that of the original languages, as three sets of criteria have emerged in the corpus analysis which underline the configuration of the translational Chinese genre system. These are (1) features related to the communicative function of translation, i.e., explicitation, simplification and interactivity; (2) source-text oriented textual and linguistic features; and (3) target-text oriented textual and linguistic features. Such corpus findings suggest that translation is a highly purposed and complex system. If we consider translational textual features like explicitation, simplification and interactivity as essentially target-audience oriented translation strategies and tactics, the corpus-driven analysis in Chapter 2 seems to suggest that the contemporary Chinese translational genre system is overwhelmingly oriented towards the target language and culture.

Chapter 3 offers an overview of translation stylistics, an important area of corpus translation research. It deploys descriptive analyses widely used in corpus-based translation studies such as the type-token ratio, standardized sentence length variation and normalized word frequency lists to explore contrastive stylistic profiles of different target versions of a source text (the case study used is from two early English translations of the Chinese literary classic *Dream of the Red Chamber* or *Hongloumeng*). The methodological framework of Chapter 3 is distinct from that of Chapter 2 in that the frequency-based analysis used in Chapter 3 is largely descriptive, whereas the quantitative methods used in Chapter 2 are more exploratory, aiming to construct new analytical instruments to make necessary preparations for further theoretical development. If we could consider the type of corpus translation research exemplified by Chapter 2 as essentially corpus-driven, the focus of the analytical strategies of Chapter 3 is to detect differences between paired translations and the source text. An important observation made in Chapter 3 regards the further analysis of the corpus findings at a social and cultural level; in other words, how to interpret the stylistic differences identified between different translations within the larger target social and cultural background – a methodological concern which points to the strengths and limitations of many similar studies on translation stylistics.

Chapters 4 and 5 reflect the theoretical branch of translation studies, which focuses on general tendencies in translations. These chapters offer two corpus-based studies of universally existent tendencies in translation, i.e. translation universals, which represent the main focus of corpus-oriented descriptive translation research. The study tests the previously untested Gravitational Pull Hypothesis (Halverson 2003, 2007, 2009, 2010). Since the Gravitational Pull Hypothesis intends to reconcile two seemingly opposing translation tendencies, full testing of this hypothesis entails testing of the mutually exclusive Over-representation of Target-Language Specific Features Hypothesis (Baker 1993, 1996) and the Unique Items Hypothesis (Tirkkonen-Condit 2001, 2004). Consequently, all three hypotheses posited on the suggested

translation universal “over- or under-representation of target-language specific features” in translation studies are tested. In order to test these hypotheses, two comparable parallel corpora having the same target language, but different source languages are needed. The feature to be tested must be unique to the target language in one of the languages pairs but must have a grammatical counterpart in the source language in the other language pair.

As a typical corpus-based study, Chapter 4 presents the design of the study, outlines the three hypotheses, the language pairs and the corpora used, as well as the grammatical structure that is tested in the case studies. In addition, Chapter 4 presents the first case study where the mutually exclusive Unique Items and Over-representation of Target-Language Specific Features hypotheses are tested. The Spanish gerund is used as the test object. In order to establish empirically that the Spanish gerund in fact constitutes a unique item in relation to Norwegian, a comparative study of 20 per cent of all of the Spanish gerunds in each text of the Norwegian-Spanish Parallel Corpus and their Norwegian counterparts is conducted.

Chapter 5 builds on the results from Chapter 4 in order to test the Gravitational Pull Hypothesis on the language pairs English-Spanish and Norwegian-Spanish, using the same grammatical structure (the Spanish gerund). The work presented in Chapters 4 and 5 demonstrates that the Gravitational Pull Hypothesis can be empirically tested using corpus data, and that the five core predictions of this hypothesis received support. In addition, the Unique Items Hypothesis was not upheld in translations from Norwegian, both with regards to frequent and to prototypical gerunds, and this raises important questions as to when this latter hypothesis applies and when it does not, and whether it is needed at all.

Since its inception in the 1980s, CTS has been one of the fastest growing research and teaching areas in translation studies as an independent academic discipline. The development of CTS owes much to the growing sophistication and specificity of related research methodologies. The current volume highlights three key research paradigms or sets of analytical strategies widely used in CTS: corpus-driven (statistical machine translation; exploratory corpus statistics), corpus-assisted (translation stylistics and parallel corpus comparison) and corpus-based (translation universal features or general translation tendencies) approaches. As the case studies used in each chapter demonstrate, each approach has its strengths and limitations, which reflects the very nature of empirical translation research. The delimitation of these three sets of distinct yet related research schemes contribute to the further expansion of the field, which relies to a large extent on the development of a robust, integrative, and innovative methodological system for empirical translation research.