

1 Advances in Empirical Translation Studies

Meng Ji

1.1 Advances in Product and Process-Oriented Translation Studies

This emerging academic discipline was divided into pure and applied research schemes in the Holmes-Toury map of translation studies (Toury, 1995). Within the pure research paradigm, the two large research schemes were descriptive translation studies and theoretical translation research. These two schemes interacted with each other in the study of translation products, processes and functions as the research objects of the discipline. The proposition that descriptive translation was a key component of the field was intrinsically linked with the perception of translation studies as an empirical or scientific research field, as ‘no empirical science may make a claim for completeness, hence be regarded a (relatively) autonomous discipline, unless it has developed a descriptive branch’ (Toury, 1982). Since its inception, the mission of descriptive translation studies has been to ‘study, describe (to which certain philosophers of science add: predict), in a systematic and controlled way, that segment of “the real world” which it takes as its object’ (Toury, 1982). Toury’s vision succinctly captured the nature and significance of the empirical branch of translation studies. Perhaps more importantly, it emphasised its identity and instrumental role in transforming and advancing translation studies as a young and rapidly growing academic discipline at that time. In the four decades since Toury’s visionary statement, descriptive translation studies or its later form, empirical translation studies, has become one of the most dynamic research fields of translation studies, distinguishing itself for its constant pursuit of scientifically rigorous research methodologies to advance our understanding of translation.

In the development of descriptive translation studies since the 1990s, two of its main subfields, product- and process-oriented translation research, have emerged as productive and influential research areas which have largely driven the growth of the field. The study of translation products is chiefly text-based, and involves the contrastive linguistic analysis of source and target text pairs or comparison amongst various versions of translation with reference to the source text. Contrastive analysis between the source and target texts represents one of the

most-practised approaches to translation studies, an area of study which was informed by the search for linguistic and textual equivalence (Nida, 1964; Koller, 1995). Since the late 1990s, there has been a gradual, yet decisive shift from source and target text-pair comparison towards the search for features and recurrent patterns in translations (Baker, Francis and Tognini-Bonelli, 1993; Laviosa, 1998; Olohan, 2004). The traditionally perceived authority and influence of the original material over translation outputs has been enthusiastically debated and critically challenged (Toury, 1980; Puurtinen, 1989; Lambert, 1995; Baker, 1996; Laviosa, 2004). The shift towards target-oriented translation research has benefited from the increasing availability of digital language corpora, which facilitate the testing and verification of research hypotheses. As the corpus-based approach to translation-product analysis has become widely accepted, a new wave of research innovation has crystallised in the new research field of corpus-translation studies (Baker, 1993, 1996; Tymoczko, 1998; Laviosa, 2002; Granger, 2003).

Process-oriented translation study has followed a distinct path of development. The process-oriented branch concerns the translators' behaviour and skill development (Wilss, 1996). More recent process-oriented studies have explored the cognitive mechanisms of translators and interpreters (Jakobsen, 2006). Interpreting is the oral practice of cross-cultural and cross-lingual communication, which has gained importance in recent years due to growing demands for high-quality interpreters for special purposes, such as legal, medical and conference interpreting. The process-oriented research paradigm has come to be known as *interpreting studies* in recent years (Salevsky, 1993; Pöchhacker and Shlesinger, 2002). This research paradigm has been strongly associated with and informed by advances in cognitive studies and psychology. Important and fruitful research efforts have been made to explore the cognitive mechanisms that underlie translators' or interpreters' learning patterns and working styles under natural or purposely designed experimental conditions, with a view to improving the translation or more often, the interpreting process and outcomes (Denzin, 2008).

Over the past few decades, product- and process-oriented translation paradigms have evolved into two distinct and interdisciplinary research fields: whilst the product-oriented research scheme has been chiefly applied in the study of written translations, the process-oriented scheme has been explored extensively using natural or simulated interpreting data. The two subfields have developed distinct empirical methodologies informed by different theoretical hypotheses and priorities. Text-based descriptive translation research has been strongly influenced by theoretical assumptions that seek to identify and account for universal patterns in translations across languages. Translation universals are descriptive hypotheses which have been formulated to capture the relationship between the source and target texts, as well as the relations between

translations and original texts written in the target languages. Translation universals include, for example, the tendency to simplify, normalise or conventionalise source language and textual patterns when translating into a distinct language and cultural system (Mauranen and Kujamäki, 2004). In terms of the differences between translations and untranslated comparable texts in the target language, the concept of translationese has been created to describe any consistent patterns or salient features of translations as compared with texts written in the original target language (Tirkkonen-Condit, 2002). Unlike translation universals, translationese is more language-dependent and provides a useful and practical framework with which to gauge the impact of the source text on the target language. Translationese can be analysed productively at morphological, lexical, syntactical, phrasal, lexico-grammatical or phraseological and typographical levels.

By contrast, process-oriented research has been informed by theoretical assumptions from cognitive or neuroscience and psycholinguistics (Flores D'Arcais, 1978; Chernov, 1979). Despite their distinct research topics and foci on translation or interpreting materials, these two subfields, as aspects of descriptive translation studies on the Holmes-Toury map, continue to share an explicit and strong emphasis on the analysis and modelling of empirical data in order to test and verify research hypotheses pertinent to translation phenomena (Shlesinger, 1989; Gile, 1994; Lambert and Moser-Mercer, 1994). In more recent times, with the introduction of quantitative research methodologies, both subfields have become increasingly exploratory and experimental, as important and revealing patterns have begun to emerge from the systematic processing and modelling of large amounts of translation or interpreting corpus data (Shlesinger, 1998; Tirkkonen-Condit and Jääskeläinen, 2000). Descriptive translation research is no longer confined to the objective documentation and recording of translation information, but is gradually moving towards the identification of recurrent or predictive patterns in translations, as Toury rightly envisioned four decades ago.

1.2 Increasing Interaction between Descriptive and Applied Translation Studies

Another important direction in the development of descriptive translation studies is that the field has developed a strong association with applied translation research, as the findings from 'pure' translation studies have both informed and benefited from advances in applied translation studies. The applied branch within the Holmes-Toury framework of the discipline covers practical research areas from translation training and translation aids to translation criticism. Translation aids encompass a wide range of translation tools and instruments such as multilingual glossaries, dictionaries and, more recently, digital

resources; these include language corpora, terminologies and integrated translation-resource creation and management systems such as local or cloud-based translation memories (Wright and Wright, 1993; Bowker, 2002). Translation aids, tools or information-management systems enable the construction of parallel or comparable corpora of large sizes. The development of language corpora has proven instrumental and cost-effective for both applied translation studies, such as translation training and teaching, and theoretical translation research, such as descriptive translation studies. The exploration of digital language corpora facilitates the identification, retrieval and quantitative analyses of textual and linguistic patterns in translations, which lies at the heart of descriptive translation studies. The creation and utilisation of digital language corpora has offered an important platform of cross-disciplinary collaboration, for example between translation-studies scholars and computer scientists on the development of hybrid translation models that integrate translation memories and machine translation systems (see Chapter 11). With the rise of globalisation, new research fields have emerged that can be incorporated into the applied branch have emerged, and that have significantly changed the landscape and our understanding of translation studies. These highly interdisciplinary and technology-intensive research fields examine aspects such as audiovisual or multimedia translation and localisation, speech-to-text or speech-to-speech translation, and service- or user-oriented multilingual translation applications which are illustrated by the case studies presented in this book.

Within the disciplinary framework devised by early translation studies scholars, the division between the descriptive and the applied research schemes has influenced the interaction and the cross-fertilisation of ideas and methodologies between the two research paradigms. For a long time after the circulation of the disciplinary map, descriptive research was largely driven by intellectual efforts to verify or contest theoretical assumptions such as translation universal patterns, norms and laws, whereas the applied branch continued to evolve, whilst the global translation industry thrives and diversifies amidst intensified intercultural and interlingual communication at a modern industrial scale. New applied research has emerged which provides academic support to new forms of industry-based translation and interpreting practices, such as game and web localisation, terminology standardisation and subtitling and dubbing for the global entertainment industry (O'Hagan and Ashworth, 2002; Remael and Díaz-Cintas, 2004).

The key adjectives used by early translation studies scholars to define and delineate the boundaries or the identity of the field were *theoretical*, *descriptive* and *applied*. In the early stages of its development, descriptive translation studies was intrinsically related to and influenced by theoretical translation research. In the following decades, this subfield of 'pure' translation research

became increasingly scientific, problem-oriented and data-driven, characterised by its clear and strong focus on the robustness of the analysis conducted on translation materials, and the verifiability and reliability of the outcomes obtained through replicable research designs. It is useful to notice that since the 1990s, as the two subfields of descriptive translation research have undergone rapid growth, they have developed distinct research methodologies: *corpus*, *corpus-based*, *corpus-driven*, *frequency analysis* or *pattern recognition* have become common descriptors for product-oriented research, whilst *experimental*, *empirical*, *cognitive mechanisms and functions* or *psycholinguistic modelling* emerge as recurrent keywords in process-oriented translation or interpreting projects.

In more recent times, useful and productive efforts have been made to deliberately integrate the two sets of research methodologies to advance descriptive research, especially the empirical study of interpreting data. The benefits of combining and leveraging the two sets of empirical research methodologies are evident. On the one hand, the construction and development of parallel or comparable corpora with data gathered from the monitoring of real-time interpreting assignments can enable the discovery of useful patterns that reveal the association between interpreters' cognitive capacities and features of interpretation outputs. On the other hand, the construction of translational or interpreting corpora, which entails the investment of significant time and resources, can be informed by theoretical hypotheses formulated for process-oriented research to increase the cost-effectiveness of the digital corpus resources being built. In this new wave of methodological exploration, a large and growing number of innovative interpreting research projects have been designed to engage with the corpus-based or the corpus-driven approaches, which were developed initially for product-oriented or text-based translation studies.

Less than four decades after the development of the Holmes-Toury map of translation studies, this young and once 'invisible' academic discipline has successfully developed its own disciplinary identity. It has evolved amidst waves of globalisation to encompass a wide range of highly specialised, data- or technology-intensive subfields that have important social and practical applications. As discussed above, the shared and increasing use and exploration of translational corpora and empirical resources has reduced, and continues to push, the intra-disciplinary boundaries imposed in early times between the fields of descriptive and applied translation research. Process-oriented research has contributed to the emergence and the rapid growth of the new research field of interpreting studies. The aim of this book on advances in empirical translation studies is to reflect and add momentum to this general trend of disciplinary growth and development by encouraging continued exchange, interaction and dialogue between descriptive and applied translation studies. This trend has

attracted great attention as the data-based and socially oriented turn in translation studies. In a manner differentiated from early disciplinary mapping, which separates descriptive and applied research, this book integrates these subfields based on their increasingly shared use of empirical language resources and advanced research methodologies to identify, analyse and provide practical solutions to changing social issues and research topics pertinent to translation studies.

1.3 Advances in Empirical Research Methodologies

The development of empirical research methodologies has contributed to the growth of descriptive translation studies. Since the introduction of corpus resources to translation studies in the late 1990s, there has been a constant search for scientifically rigorous and replicable methodologies with which to move the translation debate from emotional and rhetorical arguments to more data- and evidence-based research. Such evidence-based research facilitates the discovery of underlying patterns and features of translation products and processes that can be predicted, controlled and managed for better translation practice, training and education. Research efforts and discussions have revolved around topics closely related to empirical research methodologies. These include the purposes and aims of using language resources of varying sizes in empirical translation studies; the functionality and representativeness of different types of language corpora and their impact on the validity and wider applicability of corpus findings (Teubert, 1996); the role and relevance of theoretical hypotheses in the study of translation corpora; the advantages and limitations of using corpus-based (with theoretical assumptions) versus corpus-driven (without theoretical assumptions) approaches in the exploration of language corpora; the productivity of the combined use of different types of language corpora or the triangulation of corpus information; and the feasibility and reliability of using advanced quantitative methods to process large amounts of corpus data.

Early descriptive or corpus-based language studies focused on the search for meaningful patterns in language corpora (Hunston and Francis, 2000; Bowker, 2001; Baker, 2004). A number of frequency-based indicators and examples of subject-specific terminology have been developed to facilitate the analysis of corpus texts. These include type-token ratios, keywords and key-ness, low- and high-frequency words, hapax legomena, word clusters, word collocates and collocation, colligation, n-grams and so on (Kenny, 2014). Language-specific terminologies and corpus-analysis systems have been developed for character-based writing systems, for example, Chinese, which has been studied extensively at morphological and lexical levels; these developments have been useful in the study of low- and high-stroke characters, lexical density and

difficult words. Recently, more sophisticated analytical schemes such as lexical complexity and textual readability or accessibility have been enabled by means of widely used corpus-analysis software that can effectively process multilingual written scripts. These frequency-based indicators for corpus analyses represented the building blocks of early corpus translation studies, which through constant peer discussions of findings uncovered through corpus text analysis have since gradually led to the development of a set of widely tested and largely replicable empirical research methodologies. Cognitive translation studies has also made important use of empirical research methodologies in the collection, such as the testing and evaluation of translation and/or interpreting resources and data (Tirkkonen-Condit and Jääskeläinen, 2000; Shreve and Angelone, 2010; O'Brien, 2011). The empirical branch of translation studies, which is capable of studying, describing and predicting translation phenomena in a systematic and controlled way, as predicted by Toury, has taken shape.

Whilst the study of useful or meaningful patterns in corpus texts through close observation was widely practised at the early stage of corpus translation research, this soon proved less productive or reliable with the large amounts of parallel, comparable or translational corpus data that were made available for empirical translation research. There has been a growing need for more advanced and systematic analytical methods with which to delve into quantitative multilingual corpus data and to unlock the potential of the once controversial corpus-driven approach to descriptive translation studies. Efforts have been made to introduce and adapt statistical methods from cognate fields such as quantitative linguistics and textual statistics to the young research field of corpus-translation studies (Oakes and Ji, 2012). This trend represents the second stage in the methodological advancement of the field, as the introduction of inferential statistics has solved one major bottleneck issue within it, i.e. what is to follow after the identification of 'meaningful' textual patterns that are largely based on individual researchers' observation of limited amounts of corpus data.

With the introduction of inferential statistics, the study of language corpora has moved away from one-dimensional to two- or multidimensional analysis, as researchers are now able to study the complex correlation or causal relationship between different sets of linguistic and textual features retrieved from large-scale language corpora. Different sets of recurrent textual patterns and linguistic features of translations that used to be studied in isolation can now be analysed within a single statistical model to reveal any latent associations amongst them. These analytical techniques are particularly useful when studying and comparing differences and similarities between large-scale translational, parallel and comparable language corpora, as corpus statistics can effectively detect the impact and influence of source textual features on translations, as well as the strength of correlation between translations and the non-translation-related comparable texts

in the target language. Exploratory statistics have been introduced to the study of multidimensional translation creativity, as well as to that of the cognitive and contextual factors which may explain translators' working styles.

1.4 Towards a Social Turn in Empirical Translation Studies

The book revisits and challenges the traditional intra-disciplinary division between pure and applied research. Through various interdisciplinary case studies, it demonstrates that with the rapid growth of the field, empirical translation research that combines both product-oriented and applied research schemes has shown important potential in the development of data- or technology-intensive and socially oriented tools and analytical instruments with which to advance both applied and theoretical translation studies. This book envisions the growth of proposed data-based and socially oriented empirical translation studies in three main research areas: the development of the advanced research methodologies applied in translation studies, of large-scale multilingual resources and integrated multilingual analytical infrastructure and tools, and of user-oriented multilingual translation products and services. This requires substantial intra-disciplinary and interdisciplinary collaboration by experts from translation studies, statistics, social sciences, health, computer science and engineering. In this regard, this book makes useful efforts to expand the current horizons of empirical translation studies, especially those of corpus-translation research, by means of interdisciplinary collaboration with international environmental politics and cross-lingual health communication.

The case studies presented in this book represent important efforts to expand the current methodological framework of empirical translation studies, particularly the exploration of language corpora pertinent to socially oriented translation studies. The book proposes and illustrates new models of hypothesis formulation and verification which are being developed through interdisciplinary collaboration with 'remotely' related research fields such as environmental politics and public health. If the first stage of development of corpus-translation studies is pattern recognition, the second stage, the statistical exploration of textual patterns, is well illustrated by the case studies in this book. These case studies demonstrate the productivity of developing more sophisticated analytical instruments. These are based on high-quality multilingual resources that utilise the official multilingual terminologies and ontologies developed for specialist domains, for example, environmental agreements and laws and international environmental health guidelines and recommendations. The third stage of development of corpus-translation research aims to redefine our current understanding of descriptive or corpus translation studies as a subfield of pure translation research.

This book illustrates the development of analytical instruments that can be applied in the study of social phenomena and events at a cross-cultural and cross-lingual level. The first type of analytical instrument covers country-based ranking scales and systems that can be applied in the study of complex social phenomena, for example, multi-sectoral interaction amongst different social agents around the communication of environmental knowledge endorsed by international environmental agreements. The construction of a new ranking scale of multi-sectoral interaction was based on the computation of correlation scores amongst different pairs of social communication agencies such as governments, official reports, legal sources, top industries, major news and business sources, research institutions and digital media which facilitate the transmission and wide dissemination of translated concepts of environmental protection and sustainable development. The strength of this new ranking system, which had been developed using multilingual domain terminologies, was tested in terms of the wide range of countries that the scale can effectively compare and rank: fifteen countries in the Asia Pacific, Latin America, Europe and North America regions, with distinct social, economic and cultural profiles, were compared using the ranking scale. The validity of the multi-sectoral interaction scale was further tested by comparing the scale with authoritative international ranking systems developed for the ranking of country-based environmental performance, and the results of this showed the two ranking scales were largely consistent.

The second type of analytical instrument presented in this book is the structural equation modelling of a diffusion mechanism which imparts the attribution of social accountability across industrial sectors in terms of maintaining drinking-water safety for public health. The construction of the diffusion model utilised high-quality bilingual Japanese and English terminologies which had been developed by authoritative agencies for the translation of World Health Organization drinking-water-quality guidelines. The diffusion model tested the total, direct and indirect effects of first-tier social communication agencies on the attribution of social accountability across industrial sectors in terms of their adaptation to and compliance with international health guidelines and recommendations. The model tested the role of the digital media in enhancing or diminishing the intended effects of the efforts of the first-tier social communicators on the industrial sectors engaged in relevant social activities.

The case studies in this book illustrate recent developments in empirical translation studies, especially its increasing engagement with the study of complex and pressing social issues and problems which have emerged amidst growing multilingualism and multicultural communication in different regions of the world. This book reviews and demonstrates the increasing association between empirical translation studies, which was once categorised as 'pure' translation studies, with more applied and socially oriented translation research.

For example, the use of multilingual translation terminologies and resources can effectively assist with the development of evidence-based analytical instruments for assessing and comparing national performance in the communication of environmental knowledge; and the study of the dissemination of the translation of international health policies and guidelines can offer valuable insights into the social mechanisms that support the adaptation of global health recommendations in distinct national contexts. With the availability and convenience afforded by the increase in digital language resources and advanced corpus research methodologies, empirical translation studies has shown important potential to reveal complex mechanisms of cross-cultural and cross-lingual communication in our contemporary world. The field can thus make original contributions with which to identify and solve complex and pressing practical problems and to drive social innovation, through intra- and interdisciplinary collaboration. As the empirical branch of general translation research has advanced to its fourth stage, that of disciplinary growth, it is beneficial and necessary to review its disciplinary identity, directions of further growth and areas for research innovation that could effectively target and solve social problems such as those associated with the environment, health and social equality.

Further Reading

- Hansen, Gyde (2002). Empirical translation studies: Process and product. Copenhagen Studies in Language, vol. 27. Copenhagen: Samfundslitteratur.
- Hansen, Gyde (2003). Controlling the process: Theoretical and methodological reflections. In Fabio Alves (ed.), *Triangulating Translation: Perspectives in Process Oriented Research*, vol. 45. Amsterdam: John Benjamins, 25–42.
- Holmes, James (1975/1988). The name and nature of translation studies. In James S. Holmes, *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi, 66–80.
- Pym, Anthony, Miriam Shlesinger and Daniel Simeoni (2008). Beyond Descriptive Translation Studies: Investigations in Homage to Gideon Toury, vol. 75, Amsterdam: John Benjamins.
- Schaffner, Christina (1998). The concept of norms in translation studies. *Current Issues in Language and Society* 5(1–2), 1–9.
- Toury, Gideon (2000). The nature and role of norms in translation. In Lawrence Venuti (ed.), *The Translation Studies Reader*, vol. 2. London: Routledge.

References

- Baker, Mona (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis and E. Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins.
- Baker, Mona (1996). Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam and Philadelphia: John Benjamins.