

Translation and contrastive linguistic studies at the interface of English and Chinese: Significance and implications

RICHARD XIAO and NAIKING WEI

1. Introduction

Corpora have revolutionized nearly all areas of linguistic research over the past four decades (McEnery, Xiao and Tono 2006; McEnery and Hardie 2012). Translation studies and contrastive linguistics are no exceptions. Indeed, the rapid development of bilingual parallel corpora as well as monolingual and multilingual comparable corpora since the early 1990s has been of particular relevance and crucial importance to translation studies and contrastive linguistics.

This special issue of *Corpus Linguistics and Linguistic Theory* focuses on corpus-based translation and contrastive linguistic studies involving two genetically different languages, namely English and Chinese, which we believe have formed an important interface with its unique features as a result of the mutual interaction between the two languages. This introduction will first contextualize the special issue by exploring the state of the art in using corpora in translation and contrastive linguistic studies, particularly in the context of the two languages covered, and then provide a synopsis of each article included in this volume and comment on its significance and implications for linguistic theorization.

2. Using corpora in translation and contrastive linguistic studies

Fifteen years ago, Laviosa (1998: 474) observed that “the corpus-based approach is evolving, through theoretical elaboration and empirical realisation, into a coherent, composite and rich paradigm that addresses a variety of issues pertaining to theory, description, and the practice of translation.” The convergence between corpora and translation studies has been facilitated by three factors, namely the rapid development of corpus creation and analysis techniques in general and parallel and comparable corpora in particular, the increasing interest in Descriptive Translation Studies that has resulted from the paradigm shift from the prescriptive to descriptive approach to translation studies, and the hypothesis that “translation universals” (i.e. recurrent common features of translations) can be uncovered and tested by using corpus data (cf. Xiao and Yue 2009). The convergence between corpus linguistics and Descriptive Translation Studies is a natural one because corpus linguistics, as a discipline stemming from the description of real linguistic performance, supplies Descriptive Translation Studies with a systematic method and trustworthy data. As Toury (1995: 265) observes, corpus-based approaches can reveal the “regularities of actual behaviour.”

Tymoczko (1998: 652) predicated that “Corpus Translation Studies is central to the way that Translation Studies as a discipline will remain vital and move forward.” This prediction has been confirmed over the past decade, which witnessed the establishment of Corpus-Based Translation Studies as an important and fruitful area of translation research (e.g. Kenny 2001; Bowker 2002; Laviosa 2002; Teich 2003; Mauranen and Kujamäki 2004; Olohan 2004; Anderman and Rogers 2008; Saldanha 2009; Hruzov 2010; Xiao 2010, 2012; Hu 2011; Kruger, Wallmach and Munday 2011; Wang 2012).

According to the Toury-Holmes map (Munday 2001), there are three broad types of translation studies, i.e. applied, descriptive, and theoretical. Corpus-based approaches have proved to be useful in

all of these types of translation studies. Corpora can contribute to applied translation studies in three major ways, the first of which relates to corpus-assisted translating. As Bowker (1998: 631) notes, “corpus-assisted translations are of a higher quality with respect to subject field understanding, correct term choice and idiomatic expressions.” Secondly, corpora can be used to aid translation teaching and training, because “large corpora concordancing” (LCC) can help students to develop “awareness”, “reflectiveness” and “resourcefulness”, which are said to be the skills that distinguish a translator from those unskilled amateurs (Bernardini 1997). Thirdly, corpora, especially aligned parallel corpora, are essential for the development of translation technology such as machine translation (MT) systems, and computer-aided translation (CAT) tools, translation memories (TM) and terminology banks.

Descriptive Translation Studies, with a focus on the study of translation *per se*, aims to answer the question of “why a translator translates in this way” instead of “how to translate” (Holmes 1972/1988). The target-oriented and empirical nature of corpus-based approaches is in perfect harmony with Descriptive Translation studies. The research focus of Descriptive Translation Studies on translation as a product, in particular, relies upon monolingual comparable corpora of translated and native texts of the target language as well as bilingual parallel corpora to uncover evidence to support or reject the so-called translation universal hypotheses. As Baker (1993: 243) predicted in her seminal paper, the availability of large corpora of both source and translated texts, together with the development of the corpus-based approach, would enable translation scholars to uncover the nature of translation as a mediated communicative event.

Theoretical Translation Studies aims “to establish general principles by means of which these phenomena can be explained and predicted” (Holmes 1988: 71). It elaborates principles, theories and models to explain and predict what the process of translation is, given certain conditions such as a particular pair of languages or a particular pair of texts. Unsurprisingly it is closely related to, and is often reliant on, the empirical findings produced by Descriptive Translation Studies, e.g. corpus-based findings of the recurrent features of translational language, which are often referred to as translation universals (see Xiao and Dai’s article in this volume for further discussion).

Corpora of different kinds can be used for different purposes in translation studies. For example, parallel corpora are useful in exploring how an idea in one language is conveyed in another language, thus providing indirect evidence to the study of translation processes. Corpora of this kind are indispensable for building statistical or example-based machine translation (EBMT) systems, and for the development of bilingual lexicons and translation memories. Also, parallel concordancing is a useful tool for translators.

Multilingual comparable corpora are useful in improving the translator’s subject field understanding and improving the quality of translation in terms of fluency, correct term choice and idiomatic expressions in the chosen field. They can also be used to build terminology banks.

Translational corpora provide primary evidence for product-oriented Translation Studies, and for research of translation universals. If corpora of this kind are encoded with sociolinguistic and cultural parameters, they can also be used to study the sociocultural environment of translations.

Even monolingual corpora of the source and target languages are useful in translation studies as they raise the translator’s linguistic and cultural awareness in general and provide a useful and effective reference tool for professional translators as well as translation students. They can also be used in combination with a parallel corpus to form a so-called translation evaluation corpus that helps translator trainers or critics to evaluate translations more effectively and objectively.

Corpus-based approaches have not only revitalized translation studies as noted above; they have also revived contrastive analysis. Contrastive analysis was an important part of foreign language teaching methodology following WWII and remained dominant throughout the 1960s. But it lost

ground to more learner-oriented approaches such as error analysis, performance analysis, and interlanguage analysis. However, contrastive analysis revived in the 1990s, and the rapid development of corpus linguistics has been recognized as a principal reason for its revival (cf. Salkie 2002; Xiao and McEnery 2010; Xiao 2011). Contrastive analysis has greatly benefited from corpus-based approaches because corpus linguistics is inherently comparative in that not only are corpora always “pre-eminently suited for comparative studies” (Aarts 1998), but corpus analysis techniques are also comparative in essence (Xiao and McEnery 2010: 175). Unsurprisingly, corpus-based contrastive linguistics has emerged with a wealth of methodologies and thrived, addressing a wide spectrum of cross-linguistic issues (cf. Granger 2003). Of particular interest are the studies of cross-linguistic correspondence and equivalence based on parallel and comparable corpora, which have uncovered important features of the language system and language use and cast new light on the nature of translation (e.g. Altenberg and Granger 2002).

While, as can be seen from above, corpora have tremendously benefited translation and contrastive studies, corpus-based translation and contrastive linguistic studies have also significantly expanded the scope of corpus linguistic research. This special issue of the journal is concerned with corpus-based translation and contrastive linguistic studies of English and Chinese. While contrastive linguistics and translation studies have traditionally been accepted as two separate disciplines within applied linguistics, there are many contact points between the two; and with the common corpus-based approach and the usually shared type of data (e.g. comparable and parallel corpora), corpus-based translation and contrastive linguistic studies have become even more closely interconnected (cf. Ramon Garcia 2002), as demonstrated by the articles included in this special issue.

These studies combine contrastive analysis and translation studies on the basis of comparable corpora (either multilingual or monolingual) and parallel corpora of English and Chinese, two most widely spoken world languages that differ genetically. While the decision to involve English and Chinese in the research reported in this volume was largely based on the authors’ strong languages (they are all competently bilingual in Chinese and English), the significance of the typological distance between the two languages covered in these studies cannot be underestimated. In comparison with studies of typologically related languages, translation and cross-linguistic studies of genetically distant languages such as English and Chinese can have more important theoretical implications for linguistic theorization. Studying such language pairs help us gain a better appreciation of the scale of variability in the human language system while theories and observations based on closely related language pairs can give rise to conclusions which seem certain but which, when studied in the context of a language pair such as English and Chinese, become not merely problematized afresh, but significantly more challenging to resolve (Xiao and McEnery 2010: 1), as demonstrated by the articles to be introduced as follows.

3. Articles in this special issue

This special issue includes five articles, which deal with issues in the broad area of corpus-based translation and contrastive linguistic studies involving English and Chinese, with the first three focusing primarily on translation studies and the remaining two on contrastive studies.

In the first article, Xiao and Dai first critically examine the key concepts pertaining to the translation universal (TU) hypotheses that have been posited largely on the basis of translational English, while providing a systematic account of the lexical and grammatical properties of translational Chinese on the basis of balanced comparable corpora of translational Chinese and native Chinese and English corpora. The features at the lexical level include frequency profiles, word length, keywords,

word class distribution, major function words such as pronouns and prepositions, idioms and major punctuation marks, while grammatical properties investigated cover mean sentence segment length, basic sentence structures such as passives and disposal constructions as well as unique grammatical categories in Chinese such as classifiers, aspect markers, structural auxiliaries, and modal particles. The English-based TU hypotheses are reevaluated in the face of evidence from Chinese. The study fully supports the explicitation hypothesis, proposes to reformulate the simplification hypothesis and poses serious challenges to TU hypotheses such as normalization. Evidence from translational Chinese in English-Chinese translation also leads the study to attach greater significance to the TU hypothesis of under-representation of target language unique items than in previous research that has been based on closely related language pairs. Of particular interest is perhaps the insight we can gain from the study: some English-based claims and hypotheses may not be applicable to cases where a genetically different language is involved.

The mean sentence length has often been used as an indicator of translational simplification in TU research of translational English and related European languages. Nevertheless, different results have been reported in various studies, and worse still, conflicting interpretations have been proposed of the role played by the mean sentence length in textual simplification (see the review in Xiao and Dai's article in this volume). Sentence length may be an even less reliable indicator in languages such as Chinese, which are more tolerant of run-on sentences than English and similar European languages in that complete sentences in Chinese often end with commas instead of full stops. The readability of a text in a language such as Chinese that features left-branching modification is also measured differently from English. Such cross-linguistic differences between Chinese and English motivate Wang and Qin to explore new methods that assess the readability and quality of translated Chinese texts. In their contribution to this volume, Wang and Qin propose two innovative new measures, namely the Sentence Segment Length (SS-Length) and the construction load capacity (CLC), based on a bi-directional parallel corpus involving both literary and non-literary texts in Chinese and English. These new parameters arguably exert enormous influence on the readability and quality of translation and work better than the mean sentence length commonly used in previous research. In addition, this study also uses as a reference corpus a parallel text that features the transition from classic Chinese to modern Chinese to demonstrate how the construction load capacity in Chinese has historically been expanded under the influence of English-to-Chinese translation. The diachronic dimension is further explored in Guo and Chow's study of passive constructions in Chinese to be introduced below. While Xiao and Dai's study in this volume demonstrates that the claims and conclusions based on translational English are not necessarily tenable in translational Chinese, Wang and Qin's contribution clearly shows that the method and parameters developed for studying English may not work well on Chinese. These studies unmistakably highlight the desirability and value of involving typologically distinct languages in translation and contrastive linguistic studies.

In the third paper, Xu and Li take a form-meaning integrated approach to the investigation of a special construction in Chinese, i.e. splittable compounds, and their English translations, providing probably the first ever corpus-based systematic account of the structural and semantic non-correspondences, in terms of lexico-grammar and argument structure, between Chinese splittable compounds and their translation equivalents in English, on the basis of a sizeable Chinese-to-English parallel corpus. While this study has a primary focus on translation studies, a detailed contrastive analysis is also undertaken of the morpho-syntactic and semantic patterns of split and non-split uses of the chosen representative Chinese splittable compounds and their English counterparts. While the study has provided new insights into the cross-linguistic differences between Chinese and English in their lexico-grammar and argument structure, the analytical method proposed in this study, to an extent, will

inform research on similar linguistic issues, e.g. cross-linguistic equivalents of separable verbs in German and Dutch and phrasal verbs in English.

In addition to the three papers focusing on corpus-based translation studies, this volume also includes two corpus-based contrastive linguistic studies. Wei and Li's contribution examines semantic preferences and semantic prosodies of recurrent translation pairs in Chinese and English, discussing the theoretical implications of their research findings for cross-linguistic equivalence and semantic prosody. Unlike most previous studies that select translation equivalents on the basis of a bilingual dictionary or the researcher's intuitions, the authors use a statistical measure called Mutual Correspondence to extract translation pairs from a sizeable Chinese-English parallel corpus. Their results convincingly demonstrate that it is all the more evident, in this particular contrastive linguistic context than a monolingual context, that semantic prosody only belongs to patterns of co-selection, and that a word may be associated with more than one semantic prosody, thereby calling into question the common practice in corpus linguistics of treating counter examples as exploiting so-called default prosody of a word for ironical purposes. More interestingly, it shows that, given the semantic prosody, the prosodic strength is an important indicator of the degree of equivalence and mutual translatability of a presumed translation pair. Hence, the findings of this study are not only significant for translation and contrastive linguistic studies, but they also have equally important theoretical implications for research of semantic prosody in corpus linguistics.

Finally, Guo and Chow take a composite method that combines the variationist sociolinguistic approach and the comparative method of historical linguistics in exploring synchronic variation as well as diachronic development of the *bei* passive construction in Mandarin Chinese. Their research is based on a thorough analysis of the syntactic, semantic and pragmatic constraints on passive use together with a multivariate analysis of an extensive range of variables encoded in a novel written in the mid-18th century, before extensive contact between Chinese and English started, in contrast with the two comparable corpora which are respectively composed of modern native Chinese novels and texts translated from English. This study shows that the *bei* passive has evolved diachronically, and this change is reasonably ascribable to language contact between Chinese and English.

All in all, studies reported on in this special issue embody features at the interface of English and Chinese, which can be expected to have important significance and practical implications for linguistic theorizing.

Acknowledgements

We would like to take this opportunity to thank all the anonymous peer reviewers who have kindly acted as referees for their critical comments and most incisive suggestions for revision, which have tremendously improved the articles included in this volume. On a personal level, Richard Xiao acknowledges support from the China National Social Sciences Foundation Key Project (grant ref. 10ZD&127) and the Program for New Century Excellent Talents in University (grant ref. NCET-11-0460) by the Ministry of Education, China.

References

Aarts, Jan. 1998. Introduction. In Stig Johansson and Signe Oksefjell (eds.), *Corpora and Cross-linguistic Research* (pp. ix–xiv). Amsterdam: Rodopi.

- Altenberg, Bengt and Granger, Sylviane. 2002. Recent trends in cross-linguistic lexical studies. In Bengt Altenberg and Sylviane Granger (eds.), *Lexis in Contrast* (pp. 3–48). Amsterdam: John Benjamins.
- Anderman, Gunilla and Rogers, Margaret. 2008. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters.
- Baker, Mona. 1993. Corpus linguistics and Translation Studies: Implications and applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.
- Bernardini, Silvia. 1997. A ‘trainee’ translator’s perspective on corpora. Paper presented at the conference of Corpus Use and Learning to Translate. Bertinoro, November 1997.
- Bowker, Lynn. 1998. Using specialized native-language corpora as a translation resource: A pilot study. *Meta*, 43(4). 631–651.
- Bowker, Lynne. 2002. *Computer Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- Granger, Sylviane. 2003. The corpus approach: A common way forward for Contrastive Linguistics and Translation Studies? In Sylviane Granger, Jacques Lerot and Stephanie Petch-Tyson (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (pp. 17–30). Amsterdam: Rodopi.
- Holmes, James S. 1972/1988. The name and nature of Translation Studies. In James Holmes (ed.), *Translated! Papers on Literary Translation and Translation Studies* (2nd edn., 1st edn. in 1972; pp. 66–80). Amsterdam: Rodopi.
- Hruzov, Ivana. 2010. *Explicitness and Implicitness of Binary Coherence Relations: A Research Based on an Analysis of a Parallel and Comparable Corpus of Non-literary Texts and Translations*. Saarbrücken: Lap Lambert.
- Hu, Kaibao. 2011. *Yuliaoku Fanyixue Gailun* (Introduction to Corpus-based Translation Studies). Shanghai: Shanghai Jiao tong University Press.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester: St. Jerome Publishing.
- Kruger, Alet, Wallmach, Kim and Munday, Jeremy. 2011. *Corpus-based Translation Studies: Research and Applications*. London: Continuum.
- Laviosa, Sara. 1998. The corpus-based approach: a new paradigm in Translation Studies. *Meta*, 43(4). 557–570.
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.
- Mauranen, Anna and Kujamäki, Pekka. 2004. *Translation Universals: Do They Exist?* Amsterdam: John Benjamins.
- McEnery, Tony and Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, Tony, Xiao, Richard and Tono, Yukio. 2006 *Corpus-based Language Studies: An Advanced Resource Book*. London and New York: Routledge.
- Munday, Jeremy. 2001. *Introducing Translation Studies: Theories and applications*. London: Routledge.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.
- Ramon Garcia, Noelia. 2002. Contrastive linguistics and translation studies interconnected: The corpus-based approach. *Linguistica Antverpiensia*, 1. 393–406.
- Saldanha, Gabriela. 2009 *Using Corpora in Translation Studies*. London: Continuum.

- Salki, Raphael. 2002. How can linguists profit from parallel corpora? In Lars Borin (ed.), *Parallel Corpora, Parallel Worlds* (pp. 93–109). Amsterdam: Rodopi.
- Teich, Elke. 2003. *Cross-linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: De Gruyter.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
- Tymoczko, Maria. 1998. Computerized corpora and the future of translation studies. *Meta*, 43(4). 652–660.
- Wang, Kewfei. 2012. *Yuliaoku Fanyixue Tansuo* (Exploring Corpus-Based Translation Studies). Shanghai: Shanghai Jiao tong University Press.
- Xiao, Richard. 2010 *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing.
- Xiao, Richard. 2011. Word clusters and reformulation markers in Chinese and English: Implications for translation universal hypotheses. *Languages in Contrast*, 11(2). 145–171.
- Xiao, Richard. 2012. *Corpus-based Studies of Translational Chinese in English-Chinese Translation*. Shanghai: Shanghai Jiao Tong University Press.
- Xiao, Richard and McEnery, Tony. 2010. *Corpus-based Contrastive Studies of English and Chinese*. London and New York: Routledge.
- Xiao, Richard and Yue, Ming. 2009. Using corpora in Translation Studies: The state of the art. In Paul Baker (ed.), *Contemporary Corpus Linguistics* (pp. 237–262). London: Continuum.