
Lefer, Marie-Aude

ABSTRACT

Il s'agit d'un compte-rendu de l'ouvrage "Using Corpora in Contrastive and Translation Studies" de Richard Xiao

CITE THIS VERSION


DIAL is an institutional repository for the deposit and dissemination of scientific documents from UCLouvain members. Usage of this document for profit or commercial purposes is strictly prohibited. User agrees to respect copyright about this document, mainly text integrity and source mention. Full content of copyright policy is available at Copyright policy
Researchers have long called for a rapprochement between translation studies (TS) and contrastive linguistics (CL), which is made easier by the widespread use of multilingual corpora in both fields (Granger, 2003). Richard Xiao’s edited volume is an encouraging step in this direction. The 540-page book is a selection of papers presented at the international symposium on Using Corpora in Contrastive and Translation Studies that took place in China in 2008. It brings together 23 papers which, as the book’s title suggests, all address theoretical, methodological and practical issues linked to the development and use of corpora in TS and CL. A wide variety of languages and topics are dealt with in the volume. Coverage, however, is quite biased towards Chinese and the English-Chinese pair (14 papers). Comparisons are drawn between two (or three) languages, between non-translated and translated language, between translated and non-native language, and between native and learner language. Following the general introduction, in which the editor gives a short summary of each paper, the chapters are divided into three main sections, dealing in turn with corpus-based translation studies (ten chapters), parallel corpus development and bilingual lexicography (five chapters) and corpus-based contrastive studies (eight chapters). The present review describes the contributions in the order in which they appear in the volume and ends with a number of general comments. The articles are summarised briefly, due to space limitations.

The first part of the book, devoted to corpus-based TS, opens with a chapter by Marco Rocha examining the Portuguese translation of the English anaphoric ‘this’ in a parallel corpus. Relying on a typology based on grammatical function, reference type, antecedent type and antecedent phrase structure, he uncovers recurrent translation patterns with a view to implementing them in machine translation systems. In the next chapter, Yan Ding, Dirk Noël and Hans-Georg Wolf investigate the English-to-Chinese translation of fear metaphors. The study, which adopts Stefanowitsch’s (2006) Metaphorical Pattern Analysis, reveals that a number of metaphors and metaphorical expressions get “lost in translation”: the English metaphors shared by Chinese are not necessarily translated into their...
Chinese counterparts, as a result of their low degree of entrenchment in the target language. Yun Xia and Defeng Li’s chapter aims at assessing the benefits of using specialised comparable corpora of translated and non-translated language in pragmatic translation. Their case study, which relies on a small corpus of English translations of Chinese law firm advertisements and their counterparts in non-translated English, examines general text style, text informativity, internal and external focalisation as well as idiomaticity and appropriacy. In the next paper, Jun Miao and André Salem use textometrical methods to investigate translator’s notes in Fu Lei’s translation of Jean-Christophe (by Romain Rolland). This is an original contribution to corpus-based TS, as similar studies to date have not relied on corpus methods. Their analysis brings to light that Fu Lei’s notes mainly supply information about foreign personages, customs and history so as to fill cultural gaps. Ernest Wei Gao’s chapter sets out to give a cognitive account of coherence in simultaneous interpreting (SI). After theoretically introducing the Idealised Cognitive Model, which may nevertheless remain relatively obscure to readers not familiar with this framework, the author presents a corpus-based pilot study which indicates that teachers of SI should not only focus on interpreting skills, but also guide students to engage in interactive embodiment experience.

In the next chapter, Gert De Sutter and Marc Van de Velde aim to examine the factors affecting prepositional phrase placement in translated and non-translated Dutch. Their analysis of a corpus of literary texts, which is substantiated by statistical tests, shows that prepositional phrases tend to occur more frequently in the middle field (i.e. between the two verbal fields) in translated Dutch than in non-translated Dutch. Interestingly, the authors point out two possible explanations for this difference: source language influence (the novels were all translated from German, which also shows a marked preference for the middle field) and normalisation, as the middle field position is the typical position for prepositional phrases in non-translated Dutch. In the paper ‘A Parallel Corpus-based Study of Translational Chinese’, Kefei Wang and Hongwu Qin investigate the applicability of Baker’s (1993) translation universal hypotheses in translated Chinese. By examining type-token ratio, word, sentence and sentence segment length, part-of-speech distribution and compositionality in an English-Chinese bidirectional parallel corpus, their study shows that a number of translation universals do not seem to be fully supported in translational Chinese. Translation universals also lie at the basis of Richard Xiao, Lianzhen He and Ming Yue’s chapter, which investigates lexical and syntactic features of translated Chinese. After an interesting critical review of translation universal research, the authors introduce their new resource, the ZJU Corpus of Translational Chinese (ZCTC), which is the translational counterpart of the Lancaster Corpus of Mandarin Chinese. A particularly innovative feature of their approach is that they take genre variation into account.
By investigating lexical density and mean sentence length, frequency profiles, connectives and passive constructions, they present a qualified view on the validity of translation universals in translational Chinese. Importantly, the authors stress that the trends brought to light by their corpus analysis may in fact reflect English-to-Chinese contrasts, as most Chinese texts in the ZCTC were translated from English. Wang and Qin’s and Xiao et al.’s studies convincingly show the importance of extending the scope of TS, and more particularly translation universal research, to languages other than English.

The next chapter, by Federico Gaspari and Silvia Bernardini, seeks to compare non-native and translated language. This promising approach zooms in on two forms of mediated discourse to examine translation universals from a new perspective, notably by identifying a number of possible similarities between translation and interlanguage (so-called mediation universals). In their chapter, the authors introduce the new _Corpus of Non-Native and Translated English_, which so far focuses on the English-Italian language pair, and present a small-scale case study of ‘therefore’ in translated and non-native English. The preliminary results point to shared features of the two forms of mediated discourse. This innovative area of study will certainly help bring closer together research on translational English, English as a Lingua Franca and Learner English. The last chapter dealing with TS is by Defeng Li and Chunling Zhang, who review corpus-assisted translation research in China. The survey reveals that the majority of the studies under scrutiny are general reviews of the use of corpora in translation research, all of which display striking similarities in scope and content. Data-based studies of translational issues, by contrast, are scarce. The authors also deplore that translation studies in China seldom go beyond the presentation of statistical results, hence pleading for more original, empirical and rigorous data-based translation studies, in which researchers truly “make sense of the results” (p. 246).

The second part of the volume is devoted to parallel corpus development and bilingual lexicography. Kim Gerdes’s paper deals with poverty-driven bilingual alignment at the paragraph level. He explores the possibility of aligning parallel corpora without linguistic resources such as bilingual dictionaries. The main advantage of Gerdes’s alignment method, which is less ambitious than sentence alignment algorithms, is that it is not specifically designed for a given language pair and thus aims at universality. In the next chapter, Samat Mamitimin and Umar Dawut present a progress report on the Chinese-Uyghur parallel corpus project (Uyghur is a Turkic language spoken by about 10 million Uyghur people in China). The corpus contains different genres (fiction, news, science, government documents, legal texts, daily conversation), is annotated and sentence-aligned and is developed for use in Chinese-Uyghur translation and contrastive studies, bilingual lexicography and machine translation. Also in connection with the theme of cor-
pus building, Chong Zhu reports on the development of the *Multi-Media Subtitle Corpus* (MMSC). As rightly pointed out by the author, film and television subtitles have been relatively neglected in parallel corpus compilation projects. Zhu’s paper is devoted to the overall design and structure of the MMSC and to the description of the different processing phases. Sumithra Velupillai, Martin Hassel and Hercules Dalianis’s paper looks at the automatic identification of parallel and non-parallel text pairs in bilingual corpora compiled from multilingual websites, with a view to facilitating automatic dictionary construction. The study describes three experiments which aim (1) to create parallel corpora of the Nordic languages and their corresponding bilingual dictionaries from a multilingual website; and (2) to automatically identify and discard non-parallel texts in bilingual corpora. The experiments highlight the considerable potential of parallel corpus building and dictionary construction on the basis of multilingual websites. The following chapter also looks into web corpora. In their paper, Adriano Ferraresi, Silvia Bernardini, Giovanni Picci and Marco Baroni assess the relevance of large web-derived corpora in bilingual lexicography, with the aim of qualitatively evaluating these corpora. Indeed, little is known about the actual contents of these automatically compiled web resources. Their pilot study makes use of the English and French monolingual web corpora *ukWaC* and *frWaC* (consisting of 2 and 1.6 billion tokens each, respectively), and gives an account of their compilation and use to simulate the revision of the phraseological coverage in three English-to-French bilingual dictionary entries. The article shows that the two web corpora are a valuable resource for bilingual lexicographic purposes as they provide lexicographically relevant information for both languages. For English, *ukWaC* is shown to be on a par with the more costly and carefully-built *British National Corpus*, offering more up-to-date snapshots of current language usage and a better coverage of some word senses. This chapter makes an important contribution whose conclusions may be used as a basis for a new form of corpus-informed bilingual lexicography.

The third and final part of the book is devoted to corpus-based contrastive studies. It starts with Bart Defrancq’s paper, which provides an English-French contrastive analysis of ‘wh’-/qu-items in universal concessive conditionals (UCCs). While nearly all ‘wh’-items are allowed in English UCCs, only a small set of *qu*-items are allowed in French. This research topic makes it possible to examine how translators deal with lexical gaps and thereby determine which periphrastic forms are used to fill them. The study combines comparable and parallel corpus data. While the comparable corpus analysis is used to identify the ‘wh’-/qu-items that cannot be used in UCCs and to determine whether the disallowed items can be conveyed by alternative structures, the parallel corpus analysis highlights the strategies relied on by translators to fill in the gaps in French. Jianxin Wang focuses on contrastive connectors in English and Chinese and proposes a small-scale study of ‘however’
based on two parallel corpora. The preliminary results of the bidirectional analysis reveal the infrequent use of overt connectors in Chinese and show that ‘however’ is translated into as many as sixteen different Chinese connectors. In the next paper, Hui Yin conducts a corpus-based investigation into satellites in English and Chinese and shows that they are quite different in terms of their nature and frequency. While English mainly relies on verb particles that convey path (e.g. ‘up’/’down’), Chinese satellites tend to be the second elements of verb compounds and are typically motion verbs (e.g. lai “come”). Satellites are also more frequent in Chinese than in English. In their chapter on the repetition patterns of rhetoric features, Guiling Niu and Huaqing Hong contrast English and Chinese advertisements. Their study makes use of a corpus of English-Chinese matching pairs of advertisements published in two newspapers in Singapore and is based on a detailed taxonomy of rhetorical repetition (at the sound, word and phrase levels). It shows that Chinese and English ads favour different repetition patterns. However, the top four repetition devices are the same in both languages. Masahiko Nose’s contribution is devoted to comparative constructions in English, Japanese and Tok Pisin, a creole language spoken in Papua New Guinea. His study is based on the New Testament and draws on a distinction between four types of comparative constructions to bring to the fore interesting contrasts between the three languages under examination. The following chapter also presents a trilingual contrastive study based on biblical texts. In their paper, Hui-Chuan Lu and Yun-Hui Chen investigate restrictive relative clauses (RRCs) in Chinese, English and Spanish with the aim of applying their findings to language pedagogy. They argue that Keenan and Comrie’s (1977) Noun Phrase Accessibility Hierarchy hypothesis does not hold for que in Spanish RRCs, ‘that’ in English RRCs and DE in Chinese RRCs. Unsurprisingly, Spanish and English share a number of similarities when set against this hierarchy, while Chinese is markedly different. Next, Carmen Dayrell investigates the frequency and lexico-grammatical patterns of sense-related verbs (i.e. verbs whose core meaning is related, e.g. ‘study’/’analyse’/’investigate’/’examine’) in abstracts written by Brazilian graduate students in comparison with abstracts of papers published in English and Portuguese. The author clearly shows that some of the differences between the abstracts produced by novice and published writers may be due to the influence of the students’ mother tongue. The final paper, by Yuechun Jiang and Zhiqing Hu, focuses on reporting in English MA theses written by Chinese learners of English and native speakers of English. It is found, for example, that Chinese learners tend to overuse integral reporting and rely on a relatively limited repertoire of reporting structures.

An important asset of the volume is that it does justice to the variety of corpora and corpus methods currently used in TS and CL. Three salient points emerge: (1) the limited availability of corpus resources, (2) the impact of genre variation and (3) the importance of statistical tests. First, a non-negligible number of papers still
rely on (very) small corpora. Noticeably also, corpus resources are hard to come by for many language pairs, a problem which becomes even more acute when three languages are contrasted. The lack of large multilingual corpus resources explains the crucial importance of web corpora, be they derived from monolingual or multilingual websites (see Ferraresi et al. and Velupillai et al.). Second, a welcome development in TS and CL today is that a number of studies, such as Xiao et al. or Gaspari and Bernardini in the volume, now take genre variation on board. Such studies convincingly warn against drawing hasty general conclusions on the basis of mixed-bag corpora, as there is considerable variation across registers and genres within languages. Third, statistical tests, now a major concern in corpus linguistics, are being increasingly used in TS and CL (see e.g. De Sutter and Van de Velde). This is a necessary step towards bringing more scientific rigour to cross-linguistic research.

Another merit of the volume is that it clearly illustrates the possible concrete applications of TS and CL in lexicography, translator training, foreign language teaching/learning and machine translation. However, more papers which genuinely straddle the two fields of TS and CL would have been welcome. Translation and contrastive studies are clearly demarcated in the book (they are divided into two separate sections) and the question of the interplay between TS and CL is hardly addressed. Also slightly regrettable is the fact that, in many chapters, references tend to be missing or are not ordered alphabetically in the reference section. This limitation notwithstanding, the volume has been carefully edited and contains very few typos. Overall, Using Corpora in Contrastive and Translation Studies makes an important contribution to the fields of TS and CL. It constitutes stimulating and thought-provoking reading for translation scholars and contrastive linguists alike, enabling them to benefit from their respective achievements.

References


