

## Book Review

# Corpus Linguistics in Chinese Contexts

Reviewed by Jiajin Xu, Beijing Foreign Studies University, Beijing, China

Linguistics in China, applied and theoretical alike, has experienced a welcome climate of corpus-informed methodology over the last couple of decades. Corpus research in China is, more probably than not, published as individual papers or theme-specific monographs and collections (e.g. on learner corpus research, corpus-based translation studies, corpus-based discourse studies, data-driven learning, etc.). The current volume is one of the very few comprehensive corpus linguistics books pertaining to Chinese language and China topics. The chapter contributors are fairly representative of the field and truly international as well.

The plural form of ‘context’ in the book title seems to be a deliberate wording of the editors to run the gamut of such topics as Chinese language *per se*, the teaching and learning of Chinese, the teaching and learning of English by Chinese EFL learners, and media coverage on China issues. The pluralised ‘context’ too entails corpus research undertaken by an ensemble of scholars both from China and overseas, and by those speaking Chinese as their mother tongue as well as a foreign language.

Despite the thematic breadth and authorship representativeness, what is, unfortunately, missing from the complete picture of ‘corpus linguistics in Chinese contexts’ is corpus work done by Chinese scholars from Chinese departments and Computer Science departments at Chinese universities. They are denied international recognition largely on grounds of their English proficiency rather than their research quality. For comprehensive overviews of the under-reported corpus-based Chinese studies and Chinese natural language processing research in China, please refer to Xu (2015) and Feng (2006).

In the remainder of this review, detailed comments will be provided following respective chapter synopses.

The present book, *Corpus Linguistics in Chinese Contexts*, opens up with a foreword by Naixing Wei, president of the Corpus Linguistics Society of China, a preface by the editors, and a lengthy introduction by Wenzhong Li and Simon Smith which serves as an overview of the corpus research backdrop in China, which of course presents an account of chapter key points at its end.

In Wei’s foreword, he avails himself of the opportunity to be a tone-setter and prioritises once again, as he maintains elsewhere (Wei, 2011; Wei & Lu, 2014), theoretical orientations of corpus research to methodological and technological innovations, let alone pedagogical ones. His answer to the question ‘[d]o corpus linguists need to be theoretically-minded?’ (p. xiii) is unwaveringly affirmative and straightforward. Nevertheless, the disappointing fact is that eight out of nine articles in the collection do not work towards a theoretical statement of language, Hoey and Shao’s chapter being the sole exception.

The editors' preface rewinds time to the earlier *Corpus Technologies and Applied Linguistics* (CTAL-2012) international conference held at Xi'an Jiaotong-Liverpool University whose keynote addresses and hand-picked individual papers have become the basis of the current chapter book. In the short preface, one of their claims about the first ever major international corpus linguistics conference in China is not true, but it is safe, with pride, to say that it is the first corpus linguistics conference held in China which has yielded an edited volume in English published by a renowned international publishing house.

Li and Smith's Introduction then chronicles thoroughly the Chinese and English corpora based studies in China, and pushes the boundary of corpus research back to the compilation of encyclopaedias and concordances to Chinese classics in ancient China. Those gigantic historical undertakings do share many facets with the present-day corpus work in the sense that extremely large collections of texts were archived following well-defined categorical frameworks. However, the text amalgamation projects were primarily motivated for literary and religious classics exegesis, or simply as a library building endeavour, not in the least for (applied) linguistics purposes. Moreover, quantitative methodology did not have a role to play in the work. The absence of quantification in data analysis might be a practical criterion to disqualify non-corpus research from corpus research (Xu, 2014, p. 35). It is quite true that 'little achievement has been seen so far in the application of corpora to Chinese language analysis, and still less in Chinese language teaching' (p. 5), if we restrict corpus-based Chinese studies to the investigation of electronic texts only. However, a growing number of, pre-electronic and electronic, corpus-based Chinese studies have been carrying out for the last twenty or so years. Some new theoretical light has been shed on Chinese lexico-semantics, Chinese morphology and syntax in Wang (1983), He (1985), Zou (2001), Xiao and McNery (2004), Siewierska, *et al.* (2010), to name but a few. The downside of the progress, however, is that corpus methodology has not assumed a firm centre-stage role in Chinese linguistics.

## **Chapter 1. Lexical Priming: The Odd Case of a Psycholinguistic Theory that Generates Corpus-linguistic Hypotheses for both English and Chinese (Michael Hoey and Juan Shao)**

Chapter 1 argues for the applicability of the Lexical Priming theory to Chinese. Hence, the generalisability of the theory gets strengthened to some extent with the analysis of Chinese language. In addition to the theoretical extension to a different language, Chinese character based lexical priming analysis (e.g. Chinese word 好 collocating with the suffix -法) informs the collocational profiling of English morphemes, *-eries*, *-ology* and *-ism*. In the Chinese writing system, the minimal orthographic unit is 'character', which is in all likelihood the combination of phonetic and structural components. Characters consist of free-standing and bound ones, implying *prima facie* resemblance to English free and bound morphemes. Likewise, some English word components, i.e. bound morphemes, are found to collocate with and form certain semantic associations. The word component *-eries* 'is associated with CRIME and SIN on the one hand (*chicaneries, adulteries, forgeries*) and PRODUCT CREATION on the other (*breweries, bakeries, distilleries*)' (p. 27). The cases of *-ology* and *-ism* consolidate the word-component lexical priming of English. Two points are worth mentioning regarding the chapter: firstly, of the six claims of the Lexical Priming theory, the first two are fundamental ones, encapsulating the accumulative and productive nature of the lexical priming mechanism; secondly, amongst the discussion on the Lexical Priming theory in general and the examples of English lexical priming in particular, the chapter, understandably, recycles some of the previous publications of the authors concerning the Lexical Priming theory, such as the 'red wine' case, which might make it uninteresting reading for informed readers.

## **Chapter 2. Contrastive Corpus Linguistics: Cross-linguistic Contrast of English and Chinese (Richard Xiao)**

The second chapter was contributed by one of the leading Chinese corpus linguists, Richard Xiao. He reports some findings about passive constructions and classifiers based on Chinese and English corpora. He lucidly demonstrates his approach to contrastive corpus linguistics with special reference to Chinese and English. Xiao's approach starts with some heuristic lexical items, or a pool of seed words (e.g. typical classifiers discussed in the literature), and explores their syntactic contexts and genre variation. The analytical framework appears exceptionally operational and comprehensive, covering lexical, grammatical, and generic context of the linguistic item(s). Therefore, the chapter justifies itself as an exemplar of Xiao's corpus-based contrastive analysis. To be a bit picky, no coherence, however, is signified between the two cases, namely, passive constructions and classifiers. That is, why are passives and classifiers singled out as representative cases of contrastive analysis between Chinese and English? At the end of the chapter, Xiao proposes a model of Contrastive Corpus Linguistics (p. 58) blending translation studies and contrastive linguistics with interlanguage analysis. However, he fails to mention an influential similar model (i.e. integrated contrastive model) initiated some 20 years ago in Granger (1996).

## **Chapter 3. Learning Chinese with the Sketch Engine (Adam Kilgariff, Nicole Keng, Simon Smith)**

Kilgariff, Keng and Smith's chapter is a technical introduction, if not a manual, demonstrating how Word Sketch Engine is able to process Chinese data. Three main features of Word Sketch Engine, viz concordance, thesaurus, and sketch diff, are illustrated with example queries. Unique to the search of Chinese texts, both character and tokenised word searches are available, which are very considerate and helpful. Furthermore, PoS-category-based search allows for analysis at a more abstract level. For instance, measure word (a.k.a. classifier) search highlights the grammatical category searches. Sketch and sketch diff are, as always, the selling points, thereby the brand name of Sketch Engine.

## **Chapter 4. Patterned Distribution of Phraseologies within Text: The Case of Research Articles (Maocheng Liang)**

Maocheng Liang's chapter showcases a methodological innovation of a genre-informed phraseological profile across the discourse moves of applied linguistics research articles. Move-based genre analysis has been popular since its inception around early 1980s. Automatic analysis of generic features, however, has not seen much notable progress. That is to say, discourse moves are often manually-identified or hand-coded before automatic lexico-grammatical analysis can be performed. Liang's approach is a crude but effective one in the sense that applied linguistics research articles are brutally, so to speak, segmented by equal proportion, based on the users' own intuitive estimation of the sections which research articles might contain. Fortunately, applied linguistics empirical research articles are to a great extent homogeneous in terms of their overall discourse structure. The article integrates an ingenious research scheme, and a neat computational implementation of a move/segment by move/segment case study of linguistic features of applied linguistics research articles.

## **Chapter 5. Corpus' Pedagogic Processing of Phraseology for EFL Teaching: A Case of Implementation (Anping He)**

He's chapter assimilates her previous work on pedagogic processing of corpus findings, with particular reference to phraseological behaviour of EFL textbooks and other educational discourse, by her SCNU (South China Normal University) team. She explicates a three-phase model of streamlining corpus discoveries into English language teaching praxis. They are: Phase one: Identifying phrases as teaching targets; Phase two: Analysing phrase distribution and phrasal patterns; and Phase three: Transferring corpus findings to multimedia courseware. She has definitely to be honoured as a

committed pedagogy-minded corpus linguist for so many years. Her EFL corpus analyses and corpus-informed EFL courseware design and implementation are the front-running data-driven learning practices in the Chinese context.

### **Chapter 6. A Corpus Analysis of Chinese Students' (Mis-) use of Nouns at XJTLU (Wangheng Peng)**

In chapter six, Peng examines the usages of three English non-countable nouns by Chinese learners at a joint Sino-British university. In the self-compiled interlanguage English corpus of Chinese learners, such nouns as *advice*, *evidence* and *research* mistakenly bear plural markers (e.g. *advices*, *evidences*, and *researches*). Adopting a typical contrastive interlanguage analysis design, the local interlanguage corpus was compared with the native English corpus, BNC (the British National Corpus), the English as a Lingua Franca academic corpus, BAWE (the British Academic Written English corpus), and the Chinese EFL learners corpus, SWECCL (the Spoken and Written English Corpus of College Learners) to diagnose the performance of XJTLU students on noun countability. Helpful pedagogical implications are derived for Chinese EFL learners. No clear explanation as regards why the nouns were shortlisted from the analysis, however. In a contrastive interlanguage analysis of this kind, one intuitive method might be the keyword or key PoS-gram analysis instead of occasional observation of problematic nouns, if we wish to locate the major hurdles of Chinese EFL learners.

### **Chapter 7. A Corpus-based Analysis of the Use of Conjunctions in an EAP Teaching Context at a Sino-British University in China (Bin Zou and Wangheng Peng)**

Chapter seven shares the research design with chapter six, but the research focus is on an important type of cohesive ties, conjunctions. Chinese and English are in stark contrast in this regard. Conjunctions figure more prominently in English than they do in Chinese. This becomes particularly salient and conspicuous in Chinese to English translation. The under-represented conjunctions, proxies of logical relations, in Chinese texts are often recovered in the translational English texts. This stylistic preference of Chinese language somehow accounts for the overall under-representation of conjunctions in Chinese EFL learners' English essays. Even if conjunctions are used, as is the case with the general pattern of vocabulary use by Chinese EFL learners, the learners tend to cling to a limited number of high frequency 'teddy bear' (Hasselgren, 1994) conjunctions. As Zou and Peng rightly point out, the use of other conjunctions should be encouraged too (p. 154).

### **Chapter 8. Application of Corpus Analysis Methods to the Teaching of Advanced English Reading and Students' Textual Analysis Skills (Wang Haiping and Zheng Yuanyuan)**

Chapter eight is an interesting and solid study which invited EFL learners to be part of a corpus construction project. Students' awareness of text features was significantly raised. This careful experiment is definitely worthy of wider application elsewhere. However, as Xu (2009, pp. 44-45) commented in a review on the Lexical Approach—a corpus-inspired English teaching methodology, 'Does it pay to design, organise, and manage our teaching curriculum with a wholesale corpus-driven approach?' His worry is concerned with the matter of efficiency given the restricted time allocation in Chinese EFL classrooms.

### **Chapter 9. An Appraisal Analysis of Reports about Chinese Military Affairs in *The New York Times* (Zhaoyang Mei, Ren Zhang and Baixiang Yu)**

This chapter is a corpus-based media discourse analysis, focusing on reports concerning Chinese military news. The authors follow a top-down procedure, starting with the systemic functional taxonomy of interpersonal meaning—the Appraisal Theory—ushered in by James Martin. Such an analytical route presupposes that the authors are clearly conscious of their research objective, that is, the evaluative resources in this case. This is particularly suitable for a critical discourse analysis as

such. The advantage of the Appraisal Theory is that it offers an extremely fine-grained hierarchical system, thereby facilitating the mapping of lexico-grammatical items onto relevant evaluative categories. The thorny issue with such theory-driven manual annotation, however, is the lack of category identification and coding reliability.

## CONCLUSION

Thanks to the chapter contributors and the editors, a handful of first-rate research projects have been presented to readers. In the current volume, corpus linguistics is envisaged in the main as a methodology. The application of corpus resources and analytical techniques has found integral and creative expression in contrastive linguistics, English for academic purposes, interlanguage studies, the teaching of English and Chinese, and media discourse analysis. Most of the studies exhibit a clear applied linguistics focus, whereas over the last few years, corpus linguistics has seen a parallel success story in theoretical linguistic issues in China too (see Zhang & Yan, 2013; Kong, 2014, etc.). Alongside what is happening in the international corpus research community, the methodological synergy or triangulation in Chinese corpus linguistics is also gaining increasing attention and popularity.

Another issue which the chapters fall short of an earnest consideration of is the notion of English as a lingua franca (ELF), which is not unanimously accepted in applied linguistics as of yet. With ELF in mind, corpus design, comparative method in interlanguage analysis and translation studies may well be rather different. The gulf between native language and interlanguage will be bridged; the distinction between translated language and target/source original language(s) will be blurred. It can be argued with some exaggeration that corpus linguistics will thus be turned on its head, given that the superiority of native English will be dismissed. In like manner, Chinese as a lingua franca will inevitably become a linguistic concern too.

To wrap up, the publication of *Corpus Linguistics in Chinese Contexts* is one of the important attempts to catapult the Chinese corpus linguists as a group to the forefront of international corpus linguistics community. It can be expected that corpus linguistics in China promises to thrive in terms of both the amount and originality of its research in the foreseeable future, despite the fact that it is, heretofore, underdeveloped, if not unheard of, from a global perspective.

*Corpus Linguistics in Chinese Contexts*

Bin Zou, Simon Smith, and Michael Hoey, Eds.

©2015 Palgrave Macmillan

204 pp.

\$96 (£58) [Hardcover]

ISBN 978-0-230-28249-0

## REFERENCES

- Feng, Z. W. (2006). Evolution and present situation of corpus research in China. *International Journal of Corpus Linguistics*, 11(2), 173–207. doi:10.1075/ijcl.11.2.03fen
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer et al. (Eds.), *Languages in contrast: Text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. doi:10.1111/j.1473-4192.1994.tb00065.x
- He, L. S. (1985). Zuozhuan shiji jiebin duanyu weizhi de bijiao [The comparison of prepositional object phrase placement between *The Commentary of Zuo* and *The Records of the Grand Historian*] [Studies in Language and Linguistics]. *Yuyan Yanjiu*, 5(1), 57–65.
- Kong, L. (2014). Yinghan yanzhe daoxiangyu duibi yanjiu [A contrastive study of speaker-oriented expressions in English and Chinese]. *Waiyu Jiaoxue yu Yanjiu [Foreign Language Teaching and Research]*, 46(2), 190–201.
- Siewierska, A., Xu, J. J., & Xiao, R. (2010). Bang-le yi ge da mang (Offered a big helping hand): A corpus study of the splittable compounds in spoken and written Chinese. *Language Sciences*, 32(4), 464–487. doi:10.1016/j.langsci.2009.08.002
- Wang, F. Y. (1983, May). Ci de pinlu he zi de fenhua [Word frequency and character differentiation]. *Paper presented at the Second Annual Conference of Chinese Linguistics Society*, Hefei, Anhui.
- Wei, N. X. (2011). *Ciyuxue yaoyi* [Fundamentals of lexis]. Shanghai: Shanghai Foreign Language Education Press.
- Wei, N. X., & Lu, J. (2014). *Duibi duanyuxue tansuo: Laizi yuliaoku de zhengju* [Phraseology in contrast: Evidence from English-Chinese corpora]. Beijing: Foreign Language Teaching and Research Press.
- Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A corpus-based study*. Amsterdam, Philadelphia: John Benjamins Publishing Company. doi:10.1075/slcs.73
- Xu, J. J. (2009). Cihui zhongxin jiaoxuefa de jiaoji guan: Lilun suyuan yu fansi [The Lexical Approach as a communicative teaching approach: Retrospect and re-assessment]. [Foreign Language Education in China]. *Zhongguo Waiyu Jiaoyu*, 2(4), 38–45.
- Xu, J. J. (2014). Xu Jiajin tan yuliaoku yuyanxue de benti yu fangfa [Xu Jiajin's view on the nature and methodology of corpus linguistics] [Corpus Linguistics]. *Yuliaoku Yuyanxue*, 1(2), 35–44.
- Xu, J. J. (2015). Corpus-based Chinese studies: A historical review from the 1920s to the present. *Chinese Language and Discourse*, 6(2), 218–244. doi:10.1075/cld.6.2.06xu
- Zhang, L. F., & Yan, C. S. (2013). Hanyu fuza mingci duanyu de du 'de' pianhao: Zhenghe yuliaoku he renzhi yuyanxue de zhengju [One-de preference in complex NPs in Chinese: Integrating corpus and cognitive evidence] [Foreign Language Education]. *Waiyu Jiaoxue*, 34(1), 30–36.
- Zou, S. H. (2001). *Yuyong pinlu xiaoying yanjiu* [Studies in frequency effects of language use]. Beijing: The Commercial Press.

*Jiajin Xu is Professor of Linguistics at the National Research Centre for Foreign Language Education, Beijing Foreign Studies University as well as a founding member of the Corpus Linguistics Society of China (CLSC). His research interests include discourse studies, second language acquisition, contrastive linguistics and translation studies, and corpus linguistics. He obtained his PhD, specialising in Corpus-based Discourse Studies, from Beijing Foreign Studies University, China. From 2008 to 2009, he was a post-doctoral researcher in the Department of Linguistics and English Language at Lancaster University, UK, working on the ESRC-funded project "A corpus-based study of split words in Chinese: Interfacing morphology, syntax and pragmatics". He has published a number of English papers in international journals, including Corpus Linguistics and Linguistic Theory, Language Sciences, Chinese Language and Discourse and ICAME Journal, and over 30 papers in high-impact linguistics journals in China.*