Corpora have become increasingly important in the last few decades. Initially, they were widely used in language technology and lexicography. Rapid advances in computing have enabled the integration of corpora into language teaching and learning. New tools for corpus construction and consultation help researchers, translators, lexicographers, teachers and language learners collect and analyse data and use corpora more effectively. However, in China corpus methods have not yet been widely adopted. Outside the corpus linguistics community there is, to date, a lack of awareness of what corpora can offer, whether in the compilation of dictionaries, the creation of corpus-informed materials, or data-driven learning tasks and approaches for use in the classroom or self-study. *Corpus Linguistics in Chinese Contexts* aims to advance the state of the art in the use of corpora in applied linguistics and to spread awareness of and expertise in corpus use in China.

The book serves as a vehicle for nationally and internationally renowned experts to share their work on corpora and technology in language teaching and research. It seeks to make an important contribution to integrating corpora and technology in language teaching and research in China, and to appeal to a wide variety of readers, including anyone who is interested in teaching languages with the use of corpora and technology. It targets, in particular, language teachers who are in China, or whose students are Chinese, or who are learning Chinese, helping them to use corpus technologies, encouraging the students to learn more autonomously or through exposure to more authentic texts. It is hoped, too, that the book will appeal to the corpus research community internationally, as well as to educational organizations and Higher Education (HE) institutions who are integrating corpus technologies into their existing syllabuses. The four chapters which do
not directly address pedagogical issues will also be of interest to general corpus linguists in China and beyond.

These four chapters (by Hoey, Xiao, Liang, and Mei et al.) could be characterized as research oriented, while all the remaining chapters are practitioner (teacher) oriented, and we did consider adopting this distinction as the organizing principle for the book. Since our chapter contributions all focused on corpus-informed analysis of either Chinese or English, it was however finally decided that the chapters should be grouped according to language treated, with work on Chinese presented first.

The remainder of this introductory chapter gives an account of corpus-informed research on the Chinese language, followed by a brief treatment of English language corpus work in China, with special attention paid to learner corpora of English built by Chinese scholars. Finally, a short summary of each chapter is provided.

I.1 Chinese language corpora

While Biber et al. (1998: 21) attribute the first English use of a corpus to Dr Johnson in 1755, the earliest corpora or concordance books in China date back rather further. They often take the form of book compilations, or 类书 Lei Shu (Encyclopedias of Books). One of the first such publications, 皇览 Huang Lan, which appeared in about the 2nd century A. D., was effectively a concordance of various books, based on a list of carefully selected key words or terms indexed either by thematic or prosodic categories, for convenient use by the royals. This tradition of concordancing books was adopted in both official and private circles in the later imperial dynasties. Different from concordances of the Bible or classic works in the western tradition, which were basically complete concordances of a specific single book, the Chinese Lei Shu usually concordanced miscellaneous books. For example, 北堂书钞 Bei Tang Shu Chao (North Hall Book Notes), compiled by Yu Shinan by the end of the Sui Dynasty (about 600 A. D.), concordanced over 800 books; and 艺文类聚 Yi Wen Lei Ju (A Collection of Literary Texts), compiled by Ouyang Xun and his colleagues in the beginning of the Tang Dynasty (624 A. D.), concordanced 1,431 books. These compilations ensured that the contents of many of the ancient books that were subsequently lost or damaged were preserved for posterity.

The earliest systematic collection of data for grammatical study, according to Teubert and Čermáková (2004), was Quirk's Survey of English Usage, which began in the 1950s. The first use of corpora in