

# Application of Corpora to English Writing Teaching\*

Ying He

Xi'an Fanyi University  
Xi'an, China 710105

**Abstract**—This paper reviews the common mistakes in Chinese students' English writing and explores different categories of corpora and their possible application in language teaching. Based on the recent studies, the author finds that corpus-aided teaching has never been attached due importance in real teaching activities. It is suggested that future language writing teaching be conducted in three aspects: teacher-prepared exercise, corpus-aided writing process, and corpus-aided evaluation. Furthermore, the problems in corpus-aided teaching are analyzed to offer reference for further study in this field.

**Keywords**—*learner corpora; native corpora; classroom teaching*

## I. INTRODUCTION

Corpus linguistics is a unique language research discipline. With the rapid development of computer technology, empirical research methods in language research have been affirmed. The research results of corpus linguistics have had a profound impact on foreign language teaching. Large quantity of raw data is available in the corpus, and many software were developed to suit for TEFL teaching. The development of corpus linguistics has changed our view of language characteristics, and we also expect to apply corpus data and techniques in the classroom. The corpus linguistics method has had a great influence on the field of L2 writing teaching and research, and broadened the method of L2 writing research. [1] Since the 1980s, the corpus linguistic approach has been considered to have a huge impact on L2 writing teaching and research and has broad application prospects.[2][3][4][5][6] In recent years, learner dictionaries, grammar books, and teaching reference books based on corpus data have been published in great quantities. However, the application of the corpus in the classroom is far behind the studies on corpus linguistic research. Yet from the very start, corpus linguists have never questioned that corpora could be directly applied in the learning and teaching procedure in second language acquisition. Corpus-based English writing research has broad prospects in foreign language teaching, and it is necessary to strengthen corpus-driven writing teaching research.[7] Solveig Granath studied the application of corpus in language teaching and found that many students realized the usefulness of this tool, particularly for their own writing.

## II. ANALYSIS OF PROBLEMS IN CHINESE STUDENTS' ENGLISH WRITING

English writing has been attached great importance due to its wide application in various exams and the difficulties to be improved. From the perspective of English learning, through writing learners can consolidate and strengthen the knowledge of vocabulary, grammar, paragraph and essays, have a better understanding of cross-cultural differences and deftly practice different styles of writing.

Poor performance in writing may be influenced by many factors, among which the following three factors may be attributed as the main objective reasons.

First, the Chinese mode of thinking would usually have an impact on students' English writing. The key for Chinese students to learn English is to overcome the influence of Chinese thinking mode, which is especially important for students to improve their writing skills. Many students, especially those in lower grades, usually follow the writing process of forming the outline — Chinese thinking — translation process. In this process, the mode of thinking, the choice of vocabulary and the choice of sentence patterns are inevitably influenced by the Chinese language.

Second, improper collocations frequently appear in students' writing. The collocations in Chinese students' English writing sometimes are more in line with Chinese habits. For instance, Chinese people would use expressions like "Her marriage was happy." instead of "Her married life was happy.", and sentences like "He got 87 points in chemistry." instead of the sentence "He got an 87 in chemistry." Website or dictionary reference would not offer satisfactory answer to problems like these, while corpora would allow student to encounter "real language" rather than made-up examples to acquire reliable information.

Third, sentence patterns tend to be too singular. The sentence patterns used by many college students in the writing process are still those common in high school classroom. Simple sentence patterns, fixed and uniformed phrase, simple and Chinese way of thinking, are problems in Chinese students' English writing and they could be partially addressed with the aid of corpus.

## III. CORPUS APPLICABLE IN ENGLISH CLASSROOM

Up to now, numerous corpora have been developed domestically and internationally, however, through careful studies we found that most of corpora were used for

---

\*This work is supported by Shaanxi Education Scientific Planning Project. (Program no: SGH 13458).

linguistic research like error analysis, frequency analysis, comparison studies etc. instead of language teaching, especially direct exploration of corpora by students.

Corpora could be classified into different categories according to various standards. In English teaching, the most frequently discussed and studied corpora are native corpora and Learner corpora.

Native corpora, or collections of texts spoken or written by native speakers of a language, are being used for the purposes of enhancing language teaching. The freely-available online corpora, the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) are typical examples of this category. When students are affected by Chinese way of thinking or not certain about the collocations, examining frequency and context occurrence of lexical collocations in native corpora would allow students to contact real language and discover precise results and sharpen their language awareness.

Learner corpora (LC) are electronic collections of foreign or second language learners texts assembled according to explicit design criteria. [8] They usually collect data from language learners to form a special type of corpus. Mukherjee & Rohrbach consider that this kind of “localization of learner corpus compilation” is one of the most promising avenues in learner corpus research. They argue that “firstly, the focus on their own students’ output will involve many more teachers in corpus-based activities and that, secondly, the exploration of learner data by the learners themselves will motivate many more learners to reflect on their language use and thus raise their foreign language awareness”. [9] Core components of learner corpus research includes foreign language teaching (FLT), second language acquisition (SLA), linguistic theory and corpus linguistics. Learner corpora are closely linked with language learning and teaching. Learner corpora can be roughly classified into two types according to the way in which they are collected and their ensuing pedagogical: learner corpora for delayed use (DPU) and learner corpora for immediate pedagogical use (IPU). [8] The former, developed for quite a long time, have been proved efficient research tools but are not used directly used in teaching/learning process. The Longman Learners’ Corpus or the Cambridge Learners’ Corpus is good examples of this type. The latter, which have been developed more recently, gradually become part of normal classroom activities and the learners are producers and users of the corpus data at the same time. IPU corpora are usually much smaller than DPU corpora and more relevant as learners work on their own productions. Up to now, great achievements have been made in constructing IPU corpora. For instance, many Chinese Learner English Corpus (CLEC), a corpus of 1 million word collections of samples of written language from Chinese learners, was cooperatively built by Professor Shi Guichun from Guangdong University of Foreign Studies and Professor Yang Huizhong from Shanghai Jiaotong University. Spoken and Written English Corpus of Chinese Learners (SWECCCL), built by professor Wen Qiu Fang, was a corpus of 2 million word collections of samples of written and spoken language from Chinese learners. Though Granger stressed that IPU

corpora could be directly applied in language teaching, however, as learner corpora tend to contain more errors than native corpora, they have mainly been utilized to study Chinese students’ typical difficulties in learning English, to analyze the characteristics of Chinese students using English, and to explore the language characteristics of students of different levels and from different backgrounds. Those researches may be helpful in building an English teaching pedagogy with Chinese characteristics in exploring the study strategies and studying on the information of Chinese learners’ language study, yet they were seldom used directly in classroom teaching activities.

The two types of corpora have their own advantages in language teaching, yet they should be combined in teaching to achieve the best results. Teachers should try to explore the best combination of the two and organize different teaching activities in which they can be effectively applied.

#### IV. CORPUS-BASED LEARNING/TEACHING

While corpus-based research has accumulated awesome statistics, corpus-aided classroom teaching made little progress. Solveig Granath also found the problem in his teaching activity and warned that: Researchers have long seen the benefits of using corpora to enhance the description of language, the regular use of corpora in the EFL classroom is still a rare occurrence. [5]

According to the writing process, we roughly split the whole writing teaching into three segments: before-writing preparation, students’ writing and after-writing evaluation.

##### *A. Before-writing Preparation*

In this part, two major tasks should be completed to facilitate the following procedure.

The first task was to introduce students to corpus analysis through some well-prepared exercise. As in curriculum for the majority English learners in China, corpus study was not set as a separate course for students. According to the interviews with students, most students would take corpus as the last resort to refer to information they need, yet most of them recognized the reliability of the information offered by the resource. The purpose of training students how to use corpora aims to make consulting a corpus as natural as looking up an item in a dictionary or a grammar book. The introduction part, which was usually omitted in domestic research papers, was explained in great details in the paper “Who benefited from learning how to use corpora?” Solveig Granath gives three examples of teacher-prepared exercise used in EFL syntax courses to demonstrate how to help students develop required skills. Solveig edited concordances to help students identify different word classes, identify relevant hits and weed out irrelevant hits and study syntactic variation. If this training is offered in other courses, then it could be shortened or even omitted in English writing course. [5]

The second task was to prepare for the writing. This part may include many activities, such as brainstorming, sample reading and familiarizing theme-related phrases etc. The

learner corpora may be of great help in this part as some learner corpora could offer sample writing and theme-related phrases for students. Some Chinese teachers have tried to construct similar corpora to facilitate the writing process and found it useful.

### *B. Students' Writing*

Many students found this tool useful particularly in their own writing. English learners, especially advanced learners, are not certain about the authenticity of their language used in writing as the lexicon, phraseology and grammar patterns are beyond what they can obtain from dictionaries or reference books. Consulting native speaker may be the best option, yet it seems impossible for them whenever they encounter problems like these. In these cases, native corpora which offer real language from native speakers may be the best choice for them. Some student would even increase their language awareness through the analysis of concordance and the follow-up discussion in class. Some may discover the linguistic rules, which is one of the most valuable functions of native corpora.

Some Chinese teachers have analyzed the use of COCA or BNC in students' English writing. Fang Ling and Wang Xingfu analyze the methods and principles of using COCA and set some examples in searching mutual information of strings, semantic searching and genres analyses about words, etc.[10] Some teacher focus on the application of learner corpora in language teaching. Wu Ying explores the practical application of learner corpora in to a teaching model and the possible methods of building the individual learner corpus for error analysis. [11] Zhen Fengchao explores direct and indirect application of learner corpora in language teaching. [12]

### *C. After-writing Evaluation*

This evaluation part could also include many activities like in-class discussion, student-student evaluation and teachers' evaluation of students' writing and further improvement. With native corpora at disposal, teachers can organize some classroom activities like asking students to research proper results in corpora. The research results of corpus linguistics, including research results on chunks, collocation and usage, keywords, error analysis, etc., could be built into a language database, which can improve the effectiveness of language teaching through providing online help, and applying research results directly to textbook development and classroom teaching. Corpus-aided evaluation can provide "data-driven learning" for advanced learners. The teacher develops a targeted small-scale specialized corpus. Students use concordance software to explore the usage of English words by comparing with the native language corpus, so that students can directly contact the complex language phenomenon under the guidance of the teacher through the actual observation of the context to master the authentic English and develop language sensitivity.

Some activities, if assisted by proper corpus, would yield better results. Some corpus-based writing platforms have been developed, such as Experiencing English — Writing

Teaching Resource Program developed by Yang Yonglin and College English Teaching Platform developed by Ge Lingling. Those platforms usually have the several functions: first, many sample writing are incorporated into the system; second, students writing could be automatically reviewed and corrected, usually a grade would be given and some mistakes would be labeled even correct expressions would be offered. Third, students' writing would be automatically put into the system and thus a writing database is set easily. Fourth, interaction between teachers and students are available and encouraged in the system. With the aid of corpus-based writing systems, writing evaluation would be more pleasant and easy for students and teachers.

## V. PROBLEMS OF APPLYING CORPORA IN LANGUAGE TEACHING

Many teachers have realized the problems in applying corpora in language teaching courses; however, those problems seem to be beyond the capacity of a teacher or even a school.

### *A. Lack of Proper Training*

Even though researchers all agreed that corpora could be widely applied in language teaching such as lexical and phraseological structure teaching, syntax teaching or grammar teaching etc., yet teachers themselves usually lack the skills needed to use this effective tool. When corpora are available to them, they do not know how to incorporate the using of this tool into teaching activities. Some teachers at university level are also researchers in corpora teaching, however, successful research does not equally mean successful teaching. Proper training courses on how to utilize this resource to analyze data, design exercise and facilitate writing should be offered to guide teachers. Most teachers are not quite clear about the function of different categories of corpora and ways of effectively apply them in different courses. Solveig Granath suggested teacher training course should link the tasks directly relevant to classroom work and must be clearly designed to include exercise in using corpora to find out about lexicon, phraseology and grammar patterns, as well as background reading on corpora and corpus linguistics. [5]

As has mentioned above, learner corpora and native corpora would be both applied in teaching process, yet their functions and using methods may be quite different. Repeated hand-on exercises should be designed for students to be familiar to different types of corpora in writing process. All these need experienced teacher equipped with professional training in corpora, yet right now these courses are not so conveniently available for teachers.

### *B. Lack of Basic Research or Application in Real Teaching*

Corpus linguistics is usually viewed as a research method or a research perspective instead of as an independent linguistic branch, therefore many theoretical frameworks of corpus linguistics were constructed on the foundation of theoretical linguistics or applied linguistics. It is quite

necessary to introduce relevant linguistics theories when applying corpus linguistics to study concrete problems.

After collecting academic papers on corpus research, the author found that many paper would focus on research such as error analysis, use of certain types of lexical or syntax structures. Among the papers with “teaching or teachers” as key words, researchers would usually put forward suggestions, discuss inspiration for teaching or only compare the numbers getting from corpus-aided writing and non-corpus-aided writing. The author holds that the research should focus on solving practical problems in teaching, such as how to scientifically apply the resource in teaching and how to design corpus-related teaching activities.

## VI. CONCLUSION

Language learning is a complex process involving a series of processes such as forming and verifying hypotheses. The corpus provides a large amount of real data for language learning. Adopting a data-driven learning model helps to develop students’ ability to learn independently. Learners analyze the syntactic features and typical collocations of words by observing corpus. This kind of discovery learning method helps to improve students’ enthusiasm for learning a foreign language. Corpus technology and methods are of great significance in the reform of foreign language teaching mode. The corpus plays a decisive role in the future of foreign language teaching, but corpus software and technological applications need to be popularized, and teaching concepts need to be updated. Corpus linguists have a long way to go to meet the challenges of teachers and students.

## REFERENCES

- [1] Lee, D. & Swales, J. A. Corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora[J]. *English for Specific Purposes*, 2006(1): 56-75.
- [2] Hyland, K. *Second Language Writing*[M]. Cambridge: Cambridge University Press, 2003.
- [3] Bloch, J. The design of an online concordancing program for teaching about reporting verbs[J]. *Language Learning and Technology*, 2009(1): 59-78.
- [4] Conrad, S. Myth 6: Corpus-based research is too complicated to be useful for writing teachers[A]. J. M. Reid(ed.). *Writing Myths: Applying Second Language Research to Classroom Teaching*[C]. Ann Arbor: The University of Michigan Press, 2008: 115-139.
- [5] Granath, Solveig. Who benefits from learning how to use corpora? I. Aijmer, K. (ed.), *Corpora and language teaching*. Amsterdam/Philadelphia: John Benjamins, 2009: 47-66.
- [6] Stapleton, P. & Radia, P. Tech-era L2 writing: Towards a new kind of process[J]. *ELT Journal*, 2009(2): 175-183.
- [7] Wang, Lifei. English Writing Teaching and Research from the Perspective of Corpus Linguistics[A]. *Proceedings of the 5th International Conference on Teaching Researching EFL Writing in China*[C]. Beijing: Foreign Language Teaching and Research Press, 2008: 2-9(Ch).
- [8] Granger, Sylviane. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. I. Aijmer, K. (ed.), *Corpora and language teaching*. Amsterdam/Philadelphia: John Benjamins, 2009: 13-32.
- [9] Mukherjee, J. & Rohrbach, J.-M. Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. B. Kettemann & G. Marko (eds), *Planning, Printing and Gluing Corpora. Inside the Applied Corpus Linguist’s Workshop*. Frankfurt am Main: Peter Lang, 2006: 205-302.
- [10] Fang Ling, Wang Xingfu. Self-study and Classroom Use of the Corpus of Contemporary American English. *Foreign Language in China*, 2010(6): 79-84
- [11] Wu Ying, Que Ziji. The application of learner corpora in foreign language teaching and research — With the data collection of error analysis as the example. *Computer-assisted Foreign Language Education*, 2005(6): 76-80.
- [12] Zhen Fengchao, Wang Hua. Application of learner corpora to foreign language teaching: Ideas and methods. *Foreign Language World*, 2010(06): 72-77.