

Corpus Construction for ESP

Zhao-Jun Chen^{1*}, Miao Huang²

1,2 School of Humanities, Yantai Nanshan University, Shandong, 265706, China

*E-mail: chenzhaojun1980@126.com

Abstract: Corpus-based research is well placed to inform the content of ESP teaching. Based on the computer information processing technology, the corpus gathers a large number of authentic language materials from the reality so that it can quickly and fairly accurately identify important language data in large-scale language materials. The corpora or the web-as-corpus in ESP contexts or resources can provide learners with examples of the right types of texts, which are relevant to their needs. By means of corpus, it is beneficial to transform traditional teacher-centered teaching into student-centered knowledge-exploring model for ESP teaching.

Keywords: ESP; Corpus; Pedagogy; ESP corpus construction

1. LITERATURE REVIEW OF ESP CORPUS

A corpus can be defined as a collection of texts compiled according to set criteria and in ESP (English for Special Purposes), as in many other areas of language study; their use has had an enormous impact. In particular, descriptions of academic discourse derived from large corpora have been used in the production of reference materials. Corpora are often considered to have two major roles in language learning, both of which are relevant to their role in ESP: direct and indirect. [1] Their direct role is as a pedagogic tool; they are used with or by students for language learning purposes in the classroom. Their indirect use is to provide the data and research knowledge upon which pedagogical materials are based. Corpus investigation allows the observation of repeated patterns in large quantities of data and thus enables

evidence - based descriptions of academic registers to be provided. In a recent critical review of the field, Granger [2] argues that the rigorous analysis of learner corpora has an important role to play in the design of appropriate corpus-informed pedagogical materials for ESP. Another focus of corpus studies has been on describing the specific features of academic discourse and its phraseology and one of its achievements has been to reveal the characteristics of different disciplines and genres. [3] Nesi presents the case for compiling small, manually-sampled corpora, rich in contextual information, for research, teaching and learning in ESP and EAP. It is possible to mine the internet for documents on specified topics, within specified domains, and with the linguistic features associated with particular genres. [4]

2. NECESSITY OF ESP CORPUS TEACHING

There is increasing evidence from corpus studies of English to suggest which language items and processes are most likely to be encountered by language users, and which therefore may deserve more investment of time in instruction. Curriculum designers and classroom teachers need to have access to this information through better reference materials and syllabuses.

For those concerned with the teaching and learning of second or foreign languages in particular, information on the distribution of the elements and processes of a language can influence pedagogy in a number of ways. From the viewpoints of Kennedy, first it can influence the content of language teaching by affecting selection of what to teach, the sequencing of pedagogy and the weight given to items or parts of the language being taught,

thus contributing directly to the content of instruction. Secondly, through the consciousness-raising of teachers about language and language use, it can show that likelihood of occurrence, or frequency of use, is an important measure of usefulness. Corpus studies can also contribute to language teaching methodology by influencing the approach to instruction and making available techniques and procedures which encourage self-access and individualized instruction through interaction with authentic, analyzed text from a corpus database.[5]

Corpus is a huge database consisted by a great number of real-world language information. It is of computer retrieval and for specialized research use.[6] It plays a more and more important role in modern linguistic research and education with unique advantages of large capacity, corpus reality, and fast and accurate retrieval. In the development of corpus linguistics, it will be a trend of building corpus for ESP, while establishing a large, comprehensive corpus.[7]

ESP is different from general English in the following aspects: first, it has strong specialty in terms of vocabulary usage, strict standard in syntactic performance, and special logical passage structure restricted by the specific literary form. These characteristics cannot be acquired by learners quickly and efficiently in traditional teaching model, while they can be clearly identified by applying corpus. To be exact, learners can form intuitive, clear cognition in a relatively short period of time.[8]

At present, ESP teaching in China easily falls into the misunderstanding of “knowledge-oriented” and “empiricism”. First, curriculum is knowledge-oriented. English is only the medium of imparting professional knowledge. For students, they just want to clarify the concepts involved, principles, and lack the concern for language phenomenon and sensitivity to linguistic characteristics; Second, ESP teachers carry out ESP teaching on basis of personal experience. EAP teachers can only solve these two problems on the basis of analyzing

students’ needs, choosing real corpus and building ESP corpus.[9]

3. ROLE OF ESP CORPUS

(1) ESP corpus can realize and guarantee the veracity of language materials in teaching content. It can also provide real conversations, professional documents and other cases in real context professionally and practically, making the actual language use consistent aspects of lexical features, text patterns, application norms, etc in teaching content and international exchange.

(2) ESP corpus can highlight the vocabulary, syntactic and textual characteristics of specific professions, so that learners can quickly grasp the main points. In a large number of paradigmatic corpuses, learners can retrieve, read, easily understand and summarize language points and usage features such as vocabulary, syntax, and discourse structure. ESP corpus can help learners to grasp the special meaning of vocabulary and the professional usage of terms as soon as possible. With the assistance of ESP corpus, students can form comprehensive understanding of characteristics, meanings and usage rules of the vocabulary and terminology.

(3) ESP corpus is conducive to the realization of student-centered ESP teaching principles. At present, traditional ESP teaching adopts the teaching mode of “reading-interpreting-grammar explanation”, which is derived from general English education. It cannot fit well with ESP education purpose in certain aspects, and even easily erase ESP characteristics of specialty and veracity in vocabulary, specificity in inner-structure of passage, effectiveness in discourse, and authenticity in context.[8] There are obvious advantages in aspects of total amount of linguistic materials, veracity of linguistic samples, variety of linguistic data, and acceptability. The construction and use of ESP corpus can highlight qualities of ESP in vocabulary, esp. terminology, such as specialization, precision, grammaticalization, writing sentence pattern and rigorous discourse,

which can effectively promote the development of ESP teaching and related research.

In essence, the corpus approach is a “decentralized” approach, which allows students to independently discover the laws of language use, without relying on any authority (such as teachers or grammar books). It can also separate the students from the traditional “center”, so that students can be more active in learning, and form their own corpus by combining corpus with their professional characteristics.[10] Students will search corpus with questions and solve the problem by analyzing the results, which can further motivate their learning motivation and keep in mind the knowledge they have found.[11]

Zhao Lizhu quantitatively analyzes the application of small corpus in aviation English teaching and confirms that it is more likely to use verbal nominalization than general English in aviation English.[9] Corpus can help students better understand the characteristics of aviation English and improve their sensitivity to linguistic phenomena. Based on the self-built aviation English corpus, learners can search the real linguistic data for specific topics, vocabulary and grammar which can be used for classroom teaching and inspire students to use self-built corpus for autonomous learning. Fu Xiaoyun has researched the construction and application of aviation English corpus. [12] Aviation English corpus select linguistic data from aviation English books, textbooks, and internet resources home and abroad, covering aerodynamics, flight control, aircraft systems and structures, engines, aviation instrumentation, aviation meteorology, aviation medicine, navigation, airport operations, airspace management, aviation accident investigation, land-air calls and aviation business English, so on and so forth. “Aviation English corpus” has strong specificity, and it has the professional characteristics that the general corpus does not possess. It is of great significance to aviation English teaching and research, teaching material development, test bank building and English

training for aviation practitioners.

4.DIRECTIONS OF ESP CORPUS CONSTRUCTION

(1)Subject-specific corpus is built based on a specific discipline (or profession). Its foothold is the corpus derived from a certain discipline or profession, such as journalistic English corpus, legal English corpus, medical English Corpus and so on. Compared with other types of ESP corpus, subject-specific corpus is highly targeted and specialized. So far it has attracted much attention in the development of professional vocabulary, the compilation of specialized dictionaries and the auxiliary machine translation.

(2)Specialized Oral English corpus will become one of the key directions for development in the future, such as the oral English corpus in medical domain, the forensic corpus in legal field and the oral corpus in the tourism industry, so on and so forth. Compared with written English, the oral discourse can provide more linguistic information, reveal the inherent characteristics and laws in real communication, and provide empirical basis for the discourse analysis.

(3)Parallel corpora are a new starting point of comparative study of language, which can facilitate ESP translation teaching. Recent studies abroad have demonstrated the validity of parallel corpus in bilingual comparison and translation studies. [13] Wei Naixing pointed out that the parallel corpus shows more diverse semantic models and in line with the actual language use. Teachers can break through the restrictions of classroom and present learners with nearly-real context, which will surely input more fresh blood to ESP classroom teaching. [14]

5. CONCLUSIONS

In short, the construction of corpus not only builds a language resource platform for the ESP classroom teaching, but also provides a rich variety of linguistic data and strong technical support for research on current ESP classroom teaching model.

It is much more abundant and diversified than ordinary multimedia teaching resources. The unique corpus technology has unparalleled advantages in cultivating students' innovative thinking and autonomous learning, impelling ESP teaching to develop in direction of personalized learning and active and exploratory learning without constraints of time and place.

ACKNOWLEDGEMENT

This paper is supported in part by a grant from "Construction of Small-Sized Bilingual Corpora for ESP Teaching" (No: J17RA052), one of Projects for Humanities and Social Sciences of Scientific Research Development Program for Colleges and Universities in Shandong Province of China

REFERENCES

- [1]Leech, G (1997). Teaching and language corpora: A convergence. In A. Wichman, S. Fligelstone, T. McEnery and G.Knowles (eds.), Teaching and Language Corpora .London: Longman. 1 – 23.
- [2]Granger, S. (2009).The contribution of learner corpora to second language acquisition and foreign language teaching. In K. Aijmer (ed.), Corpora and Language Teaching. 13 – 32. Amsterdam: John Benjamins.
- [3]Hyland, K. (2000).Disciplinary Discourses: Social Interactions in Academic Writing .Harlow, UK: Longman.
- [4]Nesi, H. (2015). ESP corpus construction: a plea for a needs-driven approach. *ASP*, volume 68: 7-24.<http://dx.doi.org/10.4000/asp.4682>
- [5]Kennedy , G. (2000). An Introduction to Corpus Linguistics. Beijing: Foreign Language Teaching and Research Press, p281-282.
- [6]He Anping. (2004).Corpus Linguistic and English Teaching. Beijing: Foreign Language Teaching and Research Press.
- [7]Cui Weixia&Wang Junsong. (2013).The Current Situation and Development Trend of Corpus Research in China *Journal of Xi'an International Studies University*, (1), 55-58.
- [8]Xu Xin. (2013).A Brief Analysis of Application and Construction of Teaching - Applied Corpus in ESP Education. *Journal of Tianjin Academy of Educational Science*, (4), 14-16.
- [9]Zhao Lizhu. (2014).Application of Small Corpora in ESP Teaching: A Case Study of Aeronautical English Teaching. *Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition)*, (2), 89-92.
- [10]Zhao Qing. (2010).The Application of Specialized Corpus in ESP Teaching. *Journal of Chongqing University of Science and Technology (Social Sciences Edition)*, (19), 182-184.
- [11]Jiang Ting. (2005).On Construction of Forensic Parallel Corpus. *Journal of Chongqing University (Social Sciences Edition)*, (4), 94-97.
- [12]Fu Xiaoyun. (2011).Construction and Application of Aviation English Corpus.*Journal of Civil Aviation Flight University of China*, (5), 52-58.
- [13]Yi Yan&Wang Kefei. (2013).The Corresponding Research of "Dajia" Based on Parallel Corpus. *Foreign Languages and Their Teaching*, (3), 49-54.
- [14]Wei Naixing. (2011).Corpus-based Contrastive Studies of Phraseology. *Journal of Foreign Languages*, (4), 32-42.