

Construction and Application of Aircraft Manufacturing Engineering English Corpus

Xiaoying Wang

Department of Foreign Language, Nanchang Institute of Technology, Nanchang, 330044, China

Keywords: English for Specific Purposes (ESP); Aircraft Manufacturing Engineering English Corpus (AMEEC); Construction; Application

Abstract: Over the past decade, corpus-based English teaching has achieved great advance both at home and abroad. This paper attempts to design and construct aircraft manufacturing engineering English corpus by the survey of ESP teaching in the School of Aviation. It is expected to explore the applications of the corpus for specialized English in improving the ESP teaching.

1. Introduction

At present, college English teaching in China is in a critical period of transformation. Many experts once again call for the way out and the direction of development from general English to English for Specific Purposes (ESP) [1]. ESP, also known as professional English, is a practical English teaching way based on professional needs. In our country, ESP teaching is based on the specific needs of students and specific majors, such as aviation English, travel English and so on. Although ESP teaching in various universities has been explored for many years, the survey still pointed out that there are widespread problems in ESP teacher education in China and the shortage of teaching materials teachers [2]. To solve the above problems, domestic experts have turned their attention to the corpus.

The corpus refers to a large-scale electronic text library built according to certain linguistic principles, using random sampling methods to collect natural language continuous texts or discourse fragments [3]. Corpus linguistics is a new type of interdisciplinary subject that analyzes language facts from a macroscopic perspective and uses a large number of real language samples as research objects, using computer tools and probabilistic statistical models to analyze language facts from a macro perspective. The corpus method is considered to be one of the three methods of modern linguistics research because of its large amount of information storage, real data, fast and accurate retrieval. Its research field has gradually expanded from vocabulary, grammar, discourse analysis, translation to language teaching, second language acquisition, etc.

2. The Current Situation of Professional English Teaching in Aviation School

The Professional English of Aviation is a professional foundation course offered by higher education institutions and is an integral part of university ESP teaching. Its content covers a wide range of areas of expertise, including aerodynamics, flight driving, land and air calling, and aviation business English. It is designed to consolidate college English and apply its practices to specific areas of aviation to develop students' ability to use English for practical work. It is unquestionable of the importance of professional ESP teaching courses of aviation for students of this major. Then, what is the current status of professional English teaching of aviation in our university, and what problems are common and how should we solve these problems? The author uses surveys and interviews to find answers.

The author first conducted a questionnaire survey on 200 students of the aviation major of the school and obtained 180 valid questionnaires. Secondly, the author interviewed 12 professional English teachers of aviation, and the related contents included teaching management, personnel and experts. The survey covers key elements of aviation ESP teaching: students, teachers, textbooks and teaching models. Through the investigation, the author found that the current situation of the

professional English teaching of aviation in our university is not effective, and the teaching mode is backward. The teaching method is simple, the teaching materials are chaotic, and the quality is not improved. The results of the survey are shown as follows:

Table 1 Students' Satisfaction with the Professional English Teaching of Aviation

	Total	Satisfied		Neutral		Unsatisfied	
Previous Graduates	70	4	5.7%	10	14.3%	56	80%
Graduates	110	16	14.5%	20	18.2%	74	67.3%
Total	180	20	11.1%	30	16.7%	130	72.2%

Table 2 Factors Affecting the Professional English Teaching of Aviation

	Total	Autonomic Learning		Teacher		Textbook		Teaching Method	
Previous Graduates	70	11	15.7%	19	27.1%	14	20%	26	37.1%
Graduates	110	10	9%	42	38.2%	46	41.8%	12	10.9%
Total	180	21	11.7%	61	33.9%	60	33.3%	38	21.1%

3. The Construction of English Corpus for Aircraft Manufacturing Engineering

With the development of computer technology, in the construction of corpus, people tend to establish a small-scale specialized corpus with more specificity and higher degree of specialization to assist language teaching and research. Some achievements have been made in terminology search, data-driven learning (DDL), and professional dictionary compilation. In the field of corpus-assisted ESP teaching, most studies are only procedural summaries, and their applicability has not been fully discussed [4]. In the author's school, the situation is not optimistic. According to the above survey, the ESP teaching in the aviation department is not effective, and the teaching mode is backward. Totally, the current situation of the teaching is not satisfied.

In view of the above situation, the author tries to establish a professional English corpus for small aircraft manufacturing engineering to improve the current situation of ESP teaching in this major, broaden the aviation professional English teaching ideas, and establish a student self-learning network platform. At the same time, the corpus is used to collect first-hand information for aviation professional English translation, language research, textbooks and compilation of specialist dictionaries. The entire corpus platform construction process includes: corpus design, corpus collection, cleaning, processing assignment and corpus retrieval and sharing. After completing the basic design of the corpus platform, the next important steps in the construction of the corpus include: corpus collection, cleaning, labeling and assignment. The information about the corpus construction procedures is shown in detail.

3.1 Collection of Corpus

The Aircraft Manufacturing Engineering English Corpus (AMEEC) is a small specialized corpus. The corpus is initially designed as a monolingual library of 200,000 words, which is an open corpus. In the future, it may be further expanded or increased according to research or teaching practice. The vocabulary, sentences and chapters selected in this corpus are as close as possible to the profession, covering all aspects of the industry, and the content is mainly based on professional literature, including the scanning of the aircraft manufacturing engineering professional English textbooks, the industry's latest foreign journals, the downloaded ICAO organization's relevant regulations and international conventions, internal airline technical information, academic papers and e-books.

3.2 Cleaning of Corpus

The original corpus obtained by scanning or downloading often has different formats, carries a

lot of redundant or even garbled information, and needs to be further cleaned into a usable raw corpus. First, we use text editors such as Text Editor, Adobe Acrobat, and Wordsmith to unify scanned and downloaded files into plain text format. The second step is to use the “text organizer” to clean all the text, eliminating extraneous symbols such as extra spaces, punctuation, garbled characters. Finally, the clean text is manually checked one by one.

3.3 Processing of Corpus

First, we use Metadata Encoder to label the corpus text with meta-information such as title, subject, source, date, and style to help search, query, and analyze the corpus later. Secondly, the CLAWS7 code-based system developed by Lancaster University in the United Kingdom is used to carry out part-of-speech assignment for the corpus to facilitate the retrieval and analysis of the grammatical features and grammatical structure of the words in the future. Finally, based on the automatic code assignment of the software, the error code is further manually corrected according to Oxford Advanced Learner’s English-Chinese Dictionary.

3.4 Retrieval and Sharing of Corpus

The completed corpus can use specialized software to retrieve the electronic text of the corpus. For the raw corpus, the research team used the corpus retrieval software AntConc and KWIC index technology to extract the word frequency Table, and selected the high frequency vocabulary to study its word frequency, collocation and language attributes. In the familiar corpus, Wordsmith, ParaConc and CLAWS can be used for code retrieval, and the above high-frequency vocabulary is further studied in terms of words, word frequency, phrases, collocations, sentence patterns and chapters.

4. The Application of English Corpus for Aircraft Manufacturing Engineering

For a long time, the focus of college English teaching has always been in public English. The research on ESP investment is obviously insufficient. Many colleges’ professional English teaching syllabus and teaching materials are all based on the experience of college English teaching. Experts pointed out that the preparation of the syllabus should be based on empirical evidence, rather than on the basis of intuition. A large number of real language use examples in the corpus can provide scientific guidance for the design of syllabus and the preparation of related textbooks.

On the platform of the English corpus of aircraft manufacturing engineering, this study extracts different information from the corpus according to the search software, and studies the professional English in language theory and teaching application. The corpus teaching practice involves all aspects of foreign language teaching: outline and textbook writing, language testing, learner language research, teacher development, etc.

4.1 Outline Design and Textbook Writing

By searching the AMEEC corpus, we find the asymmetry distribution characteristics of the linguistic images of aviation English in the real context, such as polysemy, grammatical structure, phrase and lexical collocation. According to the asymmetry feature, the corpus word frequency statistics Table is established to distinguish high frequency words, low frequency meanings of low frequency words, low frequency words, high frequency meanings of low frequency words, high frequency grammatical structures and high frequency vocabulary combinations. On the basis of the word frequency statistics Table, the function words are eliminated by “stop-list” and the word frequency, coverage and distribution of the real words are analyzed, and the aviation professional English vocabulary is developed. This language information extracted from the real language environment can make the teaching content difficult to arrange more rationally, and it can make the design of the reading materials, exercises and test questions more realistic, and further improve the English teaching of this professional, especially in the professionalism and practicality of the outline design and textbook writing.

4.2 Development of Classroom Auxiliary Teaching Platform DDL

Data-driven learning (DDL) is a new foreign language learning method based on corpus data. Its application in classroom teaching mainly refers to teachers searching through corpus, editing index of words, guiding students to observe, generalize and summarize the phenomenon of language used from a large number of corpus data, and spontaneously discover grammatical rules, meaning expressions and pragmatic features. This new student-centered teaching model can further develop a DDL learning platform that is completely self-designed, self-monitoring, self-analyzing and self-discovering.

The English DDL learning platform for aircraft manufacturing engineering is an autonomous learning method that combines corpus and data-driven learning. The completed professional English corpus is stored in the original database in the hard disk of the server for remote access by the user, and the platform is provided with unedited raw data by the platform. Students search the corpus for the content they are learning, edit the search results, generate indexes of words, and constantly think about and understand grammar rules or apply grammar features by observing and asking questions. Throughout the learning process, students are completely dominant, from one-way listening to two-way exploration, and achieving all-round language learning through human-computer interaction.

4.3 Corpus in Aviation English Testing and Training

According to the requirements of the ICAO English Language Proficiency Test in Annex I of the International Convention on Civil Aviation, the civil aviation pilots who are engaged in international air operations and domestic special routes must reach ICAO Aviation since March 5, 2008. English level 4 or higher language level standards can be used for related business operations. To this end, the Civil Aviation Administration of China independently developed the China Civil Aviation Flight Personnel English Test System (PEPEC) and the China Civil Aviation Controller English Test System (AETS) to assess the college English ability of the flight practitioners and the communication skills of the radio call language. By China Civil Aviation Administration Flight Standards Division, 2010, the two sets of test systems are scientifically designed and functioning well, and are gradually expanding the test library capacity to meet the growing requirements of our flight crew. At present, the PEPEC test system provides test questions. The Civil Aviation Administration requires all civil aviation institutions, language training institutions and airlines to prepare test questions in strict accordance with the test syllabus issued by the Civil Aviation Administration. After examination and approval by experts, they can be put into the databases for examination.

Since the correct answer is in the corpus, a corpus can be used to automatically write language test questions [5]. We can design test questions according to the examples or contexts provided by the corpus in the AMEEC corpus, or use Power GREP and other software to automatically generate exercises and evaluate and adjust the difficulty of the questions. It is also possible to further use the corpus to analyze and evaluate the objectivity and accuracy of the test questions, to avoid the occurrence of propositions, to overcome the actual situation of the language use, and to adjust and modify in time, which is beneficial to improve the stability, safety and ease of operation of the test. In addition to providing mature test questions for the test system, in the daily training exercises, the corpus can also be used to retrieve high-frequency language projects and discover the characteristics of aviation English, and determine the difficulty and difficulty of language teaching and testing.

5. Conclusion

At present, with the development of globalization, the demand for aviation talents in various countries is getting larger and larger, and the English level of practitioners is also getting higher and higher. More and more colleges and universities have chosen to establish a small specialized corpus to improve the status of aviation professional English teaching. As Sinclair predicted in 2003, “the

momentum of large-scale corpus construction is slow, which is replaced by the rise of a large number of small corpora, and the establishment of more, professional and relatively small specialized English corpus is a major trend of the future corpus linguistic development.” [6]. The corpus of applied fields at the second UCCTS conference in 2010 was the focus of research. The meeting proposed that the genre of genre (genre) affects the research results and the corpus is not limited to the literary and non-literary divisions but also the specific genre of business, tourism, medicine and aviation. Boeing and other European and American aircraft manufacturing companies have already established their own terminology and translation memory and researched their applicability. In China, due to the professionalism of aviation technology literature and large engineering volume, the related corpus construction and application research is still in the early stage. The establishment and application of the aviation professional English corpus is an attempt in this field.

Acknowledgement

In this paper, the research was sponsored by the Education Science “Twelfth Five-Year Plan” Project of Jiangxi Province (Project No. 15YB153).

References

- [1] Cai Jigang. A Subversive Foreign Language Teaching Idea and Method. *Foreign Language Teaching Theory and Practice*, 2014, vol. 2, 1-7.
- [2] Gu Zhizhong. Analysis of the Current Situation of ESP Teaching in Colleges and Universities and Coping Strategies. *Foreign Language Teaching*, 2010, vol.5, 25-29.
- [3] Yang Huizhong. *Introduction to Corpus Linguistics*. Shanghai: Shanghai Foreign Language Education Press, 2002, 183-1997.
- [4] He Pingan. *Introduction to Corpus Assisted English Teaching*. Beijing: Foreign Language Teaching and Research Publishing, 2010, 153-157.
- [5] Leech, Geoffrey. *Teaching and Language Corpora: A Convergence*. London: Longman 1997, 111-113.
- [6] Li Wenzhong, Zhai Jianzhong, Wei Naixing. Review of the 2003 Shanghai Corpus Linguistics International Conference. *Journal of PLA University of Foreign Languages*, 2004, vol.56-59.