Advances in Economics, Business and Management Research, volume 29
International Conference on Innovations in Economic Management and Social Science (IEMSS 2017)

Research and Construction of Spoken English Graded Corpus for College English Majors

Baiping Huang
Pingxiang University, Pingxiang, Jiangxi, 337055, China

Keywords: Hierarchical corpus; spoken English; English teaching; grading standard

Abstract. This paper briefly introduces the construction of a spoken English corpus including English interactive classification, the specific process of topic selection, data description, data processing and integration. To some extent, this study makes up for the lack of domestic similar research, but also provides a reference for the construction of similar corpus. At the same time, it is necessary to make clear that the classification of oral interaction is only a preliminary attempt. Limited by the sampling area, the sample will have a certain degree of convergence in the performance. The area for the investigation, its performance may be different. Therefore, it is necessary to expand the scope of the investigation in order to develop a conventional standard of oral English interaction in the future.

1. Introduction

Oral communication and written communication are the two main channels of human communication. As far as oral English is concerned, the service industry has been paid attention to by the linguists. The former is the dialogue between the customer and the service personnel, such as shops, banks, hotels, restaurants, etc. Different spoken language styles have their own language style [1-3]. In recent years, the scope of spoken language research continues to expand to areas such as the court dialogue research and interviews etc. Some linguists are interested in business talks, TV interviews, online chat. The emergence of modern spoken corpus plays an important role in the further study of oral style, and more importantly, it brings a new teaching idea to English teaching.

Corpus has been widely used in language teaching, linguistic research, language textbook compilation and dictionary compilation [4, 5]. There are special spoken corpuses, while some of them are part of a comprehensive corpus, the former is a pure spoken corpus and the latter is a non pure spoken corpus, compared with the well-known English spoken corpus: Corpus of London Teenage Language (COLTL), the Michigan Corpus of Academic Spoken English (MICASE), Lancaster IBM Spoken English Corpus (SEC). Some of these corpuses can be downloaded on the Internet or online, while others need to be purchased. Discussion on spoken English corpus of China's English education is very little, Zhao Chen (2003) "the question adjuncts in spoken English Corpus based on the pragmatic functions of investigation" and "coal (1997) of the British National Corpus and oral English study" [6, 7]. This paper also discusses the relationship between oral English corpus and oral English teaching.

2. A spoken English corpus is an effective supplement to textbooks

From the point of view of oral English teaching, the main functions of spoken English corpus are focused on the following five aspects [8]:

(1) To make up for a single textbook content. Through the corpus, in addition to learning textbooks, students can have access to a variety of styles, which broadens their horizons and increases the content and scope of language input.

(2) To provide the most authentic and reliable language information for English teachers. Spoken English corpus undoubtedly provides a reliable source of language for oral English teaching.

(3) The content of English teaching is not based on the sense of language, but on the basis of real materials, so that the content of the study is closer to the reality of life.

(4) It is helpful to carry out task based learning activities and implement material driven language learning. Spoken English corpus provides material for inquiry learning activities.

(5) Teachers and students can work together to build a corpus of spoken English for learners.

The main functions of spoken English corpus are shown in Fig. 1.

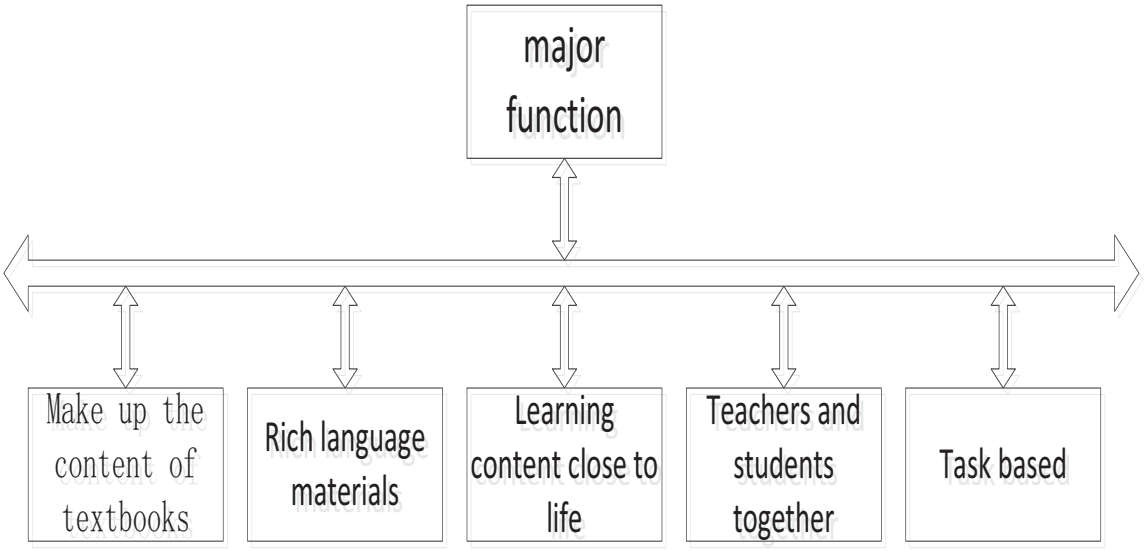

Fig. 1. Main functions of spoken English corpus

Obviously we must be aware that, against the "imperial" textbook long rules, teachers and students alike are more likely to use sources of a spoken English corpus, because the corpus language materials, unlike the "appointed" teaching materials are so "clean" and "fresh". Therefore, the concept of teachers and students need to change, only teachers and students make full play to the role of "double subject", can the quality of classroom teaching be improved.

3. Grading standard of spoken corpus

There are two levels of standards in the corpus: one is the natural classification, that is, according to the grade of the students. Because the students in each grade are about the same age, they can be classified according to their age, and the other is based on the difference of the level of oral interaction [9]. There is also a way of dividing into three levels according to age: basic, intermediate, and advanced. The basic and intermediate stage emphasizes the cultivation of knowledge ability, communication ability and intercultural communication ability and the advanced level emphasizes critical thinking.

The two types of classification can be used in the corpus to achieve free switching, which is convenient for different researchers to do different research. The data of natural classification is convenient for researchers to investigate the interaction between the students in different grades, and the difference classification is convenient for the researchers to understand the differences of the students' oral interaction ability. The Characteristics of spoken interaction is shown in Table1.

Table 1. Characteristics of spoken interaction

Interactive cooperation	Interactive steps	Topic content	Concomitant strategies
Abide by the cooperative principle and politeness principle	Direct questioning; simple response; no self correction	Familiar with the social environment; the content of the topic before and after the same, but the use of a single means	The words are simple, and the documents are simple; the words are simple, and sometimes have errors; there is no means to ease the discourse
Abide by the cooperative principle and politeness principle	Indirect questioning; response specific; self correcting	Familiar with the social environment; the content of the topic before and after a complete agreement, the use of a variety of means	The words are simple, and the documents are simple; the words are simple, and sometimes have errors; there is no means to ease the discourse

To determine the level of oral interaction classification, interactive grade is not as clear as the oral classification standards, one needs to analyze the data on the basis of summary [10-11]. We put half of the data as the observation data, and then the other half of the data to do comparison data, verify the summary of the characteristics, so as to determine the classification standard of all English spoken language interactive, which is divided into different levels with typical examples describing the features of verbal interaction.

4. Corpus Construction

In the context of foreign language learning, we believe that oral corpus is an effective supplement to the textbook. Based on the present textbooks, we need to let the students understand how native English speakers communicate in a similar environment. The comparison method can deepen the students' understanding of textbooks and spoken English corpus. The construction process of the corpus can be divided into four stages, which is shown in Fig. 2.

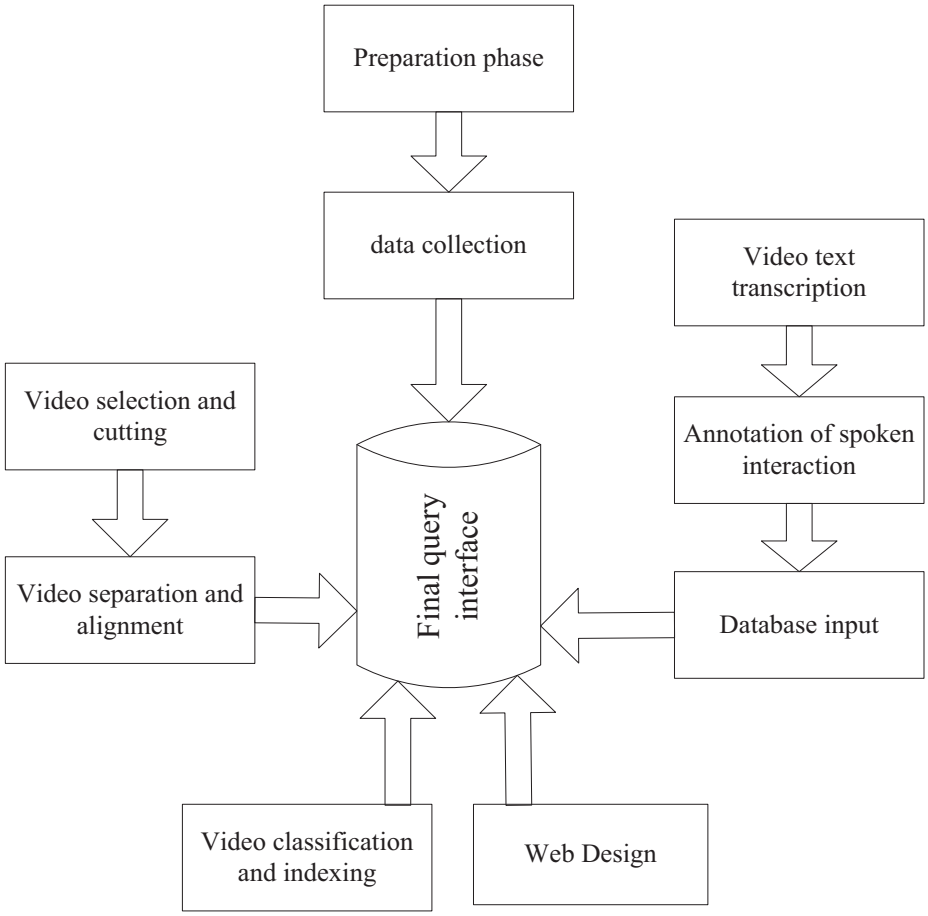

Fig. 2. The construction process of corpus

Stage One, Preparation. The main task of this stage is to carry out the work of sampling, including: literature review, to determine the level of reference points and the basic classification principles. Determine the size of the corpus, the number of samples and the length of a single. Determine test questions and conversations. Prepare related documents.

Stage Two, Data Collection Phase. The main tasks include: to enter the sample schools, according to the development of the topic of the test, according to the grade of interactive data collection, with the camera recording interactive scene. We intend to investigate the English interaction of students in 6 middle schools and colleges in Jiangxi, according to the geographical location of the distribution. Data distribution is shown in Table 2 and Fig. 3.

Table 2. Data distribution

	Survey quantity	Number of students	Male to female ratio	Sample number
Middle school	2	600	1:1	300
Vocational college	2	300	1:1	150
Regular college	2	300	1:1	150

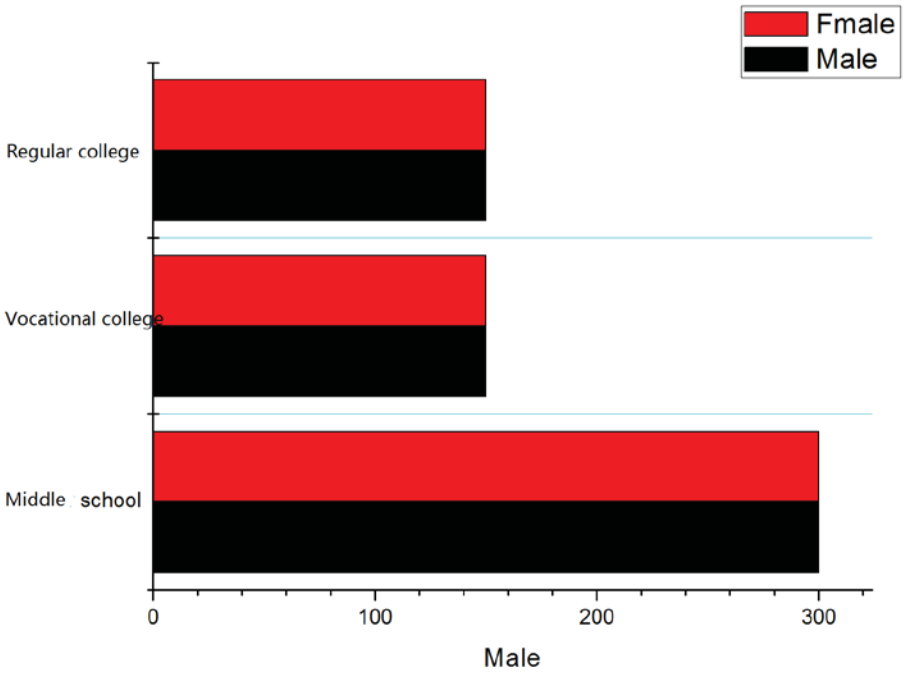

Fig. 3, Number of Students

Stage Three, Data Processing. The main steps are as follows: (1) video selection and cutting. (2) Separation and alignment of audio and video. (3) Video text transcription. (4) Tagging of oral interaction. (5) Video classification.

Stage Four, Data Integration Phase. Web design for presentation platform. Requirements can be posted online for data query function according to different users' need to achieve different processing rights. There are video, audio, texts in simultaneous visualization. Database integration, including the collation and cleaning of the text, unified format, enter the database.

In short, the construction of the corpus is of great practical significance, which provides a transliteration and labeling of the actual data for the learning of oral interactive skills. The establishment of a spoken interactive hierarchical corpus can also achieve a complete consistency of the annotation standard. The quantitative analysis of the interaction can be realized by the establishment of oral interactive graded corpus. It is helpful to the cultivation of oral interaction ability. By establishing an interactive oral classification corpus, we can understand the situation of students' oral communicative ability so as to develop reasonable teaching objective, we can also provide a reference data for the future development of a conventional grading standard. The compilation of oral interactive teaching materials will improve college English teaching.

5. Summary

The emergence of spoken English corpus provides a new platform for oral English teaching. It is of great significance to improve the efficiency of oral English teaching in Chinese colleges. This paper discusses the research and construction of a spoken English graded corpus for college English majors and characteristics of spoken corpus from the perspective of oral English teaching in China. It is one of the best measures to improve the students' autonomous learning ability. It provides a transliteration and annotation corpus of spoken interaction, based on an investigation of a quantitative analysis of oral interaction. In addition, it will also be a reference for the establishment of the classification standards and the compilation of interactive teaching materials.

Acknowledgement

This work is supported by Social Science Project of Jiangxi Province (13WX321).

References

[1] Li Y. A Corpus-Based Study on the Characteristics of "Make" in the Spoken English of Chinese English Majors[C]// 2014 International Conference on Management, Education, Business and Information Science. 2014.
[2] Wang Y. The Construction Scheme of a Graded Spoken Interaction Corpus for Mandarin Chinese [M]// Chinese Lexical Semantics. Springer International Publishing, 2016.
[3] Ling-Ling G E, Guang-Wei L I, Liu B. An Empirical Study of the Corpus-based Graded Teaching Model of College English [J]. Foreign Language & Literature, 2014.
[4] Zhu C X. A Corpus-Based Contrastive Study on Stance Markers' Using Characteristics in Non-English Majors' Spoken English [J]. Journal of Chongqing University of Technology, 2013.
[5] Peng Y M, Qu Y H. Based on Research Connecting Word Corpus of Spoken English [J]. Advanced Materials Research, 2014, 1030-1032:2689-2692.
[6] Pan C. A Cross-sectional Corpus-based Study of the Collocational Competence of the Use of ADJ+N Pattern Used by English Majors in Chinese Universities and Colleges[J]. Journal of Beijing University of Chemical Technology, 2011.
[7] Paskewitz F X. A Corpus-based Study of Recurrent Errors in the Spoken and Written English of Native Cantonese Speakers [J]. HKU Theses Online, 1999.
[8] Mair C. Writing the Corpus-based History of Spoken English: The Elusive Past of a Cleft Construction [J]. Language & Computers, 2013, 77:11-29(19).
[9] Lee J E, Kim W E, Kim K H, et al. Research on Construction of the Korean Speech Corpus in Patient with Velopharyngeal Insufficiency [J]. Korean Journal of Otorhinolaryngology-Head and Neck Surgery, 2012, 55(8):498.
[10] Zhang R. A Corpus-based Error Analysis of Students' Writing in Graded Teaching Classes [J]. Journal of Convergence Information Technology, 2013, 8(10):551-557.
[11] Wong L Y. Hong Kong's New Senior Secondary (NSS) English Language Curriculum: Perspectives from Corpus Linguistics [J]. Hong Kong Journal of Applied Linguistics, 2010.