Transactions on Computer Science and Technology
2019, Volume 7(1), PP. 24-27

Application of Corpus in Language Testing

Zhe Li

Foreign Languages Department, Changzhou Institute of Technology, Changzhou Jiangsu, 213022, China

Abstract

In recent years, the application of corpus in the field of language testing has been widely recognized, and its application potential and prospects have also attracted much attention. Based on this, first of all, the application of corpus in language testing was expounded from four aspects: test development, development of automatic scoring system, formulation of scoring scale and vocabulary formulation in examination syllabus; then, taking the construction of Chinese English learners' writing competence scale as an example, the application of corpus in the construction of writing competence scale was expounded; furthermore, the application prospects and limitations of corpus in language testing were analyzed. The rise of corpus research methods provides an effective way to improve the authenticity of language testing.

Keywords: Corpus; Language Test; Automatic Scoring System; Scoring Scale

1. INTRODUCTION

Corpus refers to a large-scale electronic library with a certain capacity, which is constructed by collecting natural and continuous language texts or discourse fragments according to certain linguistic principles and using random sampling method. In recent years, many linguists have been diligently exploring the application of corpus in language research, syllabus formulation, textbook compilation, dictionary compilation, natural language processing, machine translation and so on. It can be seen that corpus has great potential in language testing. The rapid development of corpus has contributed to the establishment of corpus linguistics, and the emergence of corpus linguistics has exerted a tremendous influence on language research. It is no exaggeration to say that corpus and corpus research have brought about a revolutionary change in language research and its application in the past decades. The corpus research methods have achieved a lot in the fields of language research and natural language processing, so the application of corpus to language testing has gradually enriched. Based on the theoretical foundation of corpus application in language testing, this paper explored the specific application of corpus in language testing.

2. APPLICATION OF CORPUS IN LANGUAGE TESTING

2.1 Test development

In test development, corpus plays an important role, especially in the selection, compilation and proofreading of test questions [1]. In the process of question-setting, the propositioner can search the source, genre, sociolinguistic variant, time or difficulty of the selected material at the same time when searching for the real material for the test questions, so as to label the native corpus quickly, compare the details of the selected material with the native corpus, and verify the representativeness and authenticity of the newly compiled test materials [2]. In the development and compilation of test questions, corpus can help the compilers to complete the corrections faster and better.

2.2 Development of automatic scoring system

The automatic scoring system is often used in the writing section. Now automatic scoring system can be divided into text category technology, latent semantic analysis technology and multi-level language feature recognition technology, the above technologies have their own characteristics and similarities [3]. The same point is that personnel can find out the linguistic features that can be referred to again after the analysis of the samples of manual

grading, such as sentence structure, text structure and writing content, and then grade other compositions. The corpus can maintain consistent scoring standards for a long time in the training of raters and the standard scoring process [4].

2.3 Development of scoring scale

When using linguistic competence to describe language, the scoring scale will be used. The corpus can be used to classify language scales and describe each stage of language competence development in order from low to high [5]. In some subjective examinations, such as oral English or writing, the rater will score the language behavior of examinee according to the scoring scale. Among the language characteristics of the candidates, good spoken scoring scale reflects the candidates' language proficiency, and it is also an important guarantee to improve the reliability between raters [6]. The corpus is of vital importance in the formulation of both writing and oral scoring scales, it can make the scoring table more authentic, objective, reliable and general.

2.4 Formulation of vocabulary in examination syllabus

When writing questions, test writers can use corpus to combine the knowledge they have learned with their own experience, and make an objective and reliable vocabulary list according to the frequency of vocabulary in the actual use of language, so as to avoid the subjectivity and arbitrariness in collecting test vocabulary to the greatest extent.

3. APPLICATION OF CORPUS IN CONSTRUCTING THE WRITING COMPETENCE SCALE FOR CHINESE ENGLISH LEARNERS

The construction of the writing competence scale mainly follows the general construction plan of the China Standards of English (CSE). Based on a large number of writing competence descriptions collected and collated in the early stage, experts, teachers and students are invited to judge the difficulty of descriptions, so as to achieve the grading verification of descriptions [7]. However, as a productive linguistic competence, writing competence has its unique advantage in a large number of collectible, preservable and observable writing texts. Moreover, Chinese scholars have established some large-scale written corpus (as shown in Table 1). We should make full use of the information provided by the corpus to make up for the problem that experts and teachers may have too much subjectivity in the process of judging, so as to supplement and improve the scale.

TABLE 1 WRITTEN CORPUS OF CHINESE-SPEAKING ENGLISH LEARNERS ESTABLISHED BY CHINESE SCHOLARS

Name	Storage Capacity	Construction Unit	Main Corpus Sources
Chinese Learner English Corpus	1 million	Guangdong University of Foreign Studies, etc.	National College English Band 4 and Band 6 Compositions
Chinese English Students' Oral and Written Corpus	1 million	Nanjing University	Written Corpus: Time-limited Composition and Extracurricular Non-time Composition of Grade 1-4 English Majors in 9 Colleges and Universities at Different Levels
Experience English Writing Database	800 thousand	Tsinghua University	Student Works of Tsinghua University
English Corpus for Middle School Students	876 thousand	South China Normal University	English Composition in College Entrance Examination

Firstly, at present, the written corpus established by Chinese scholars mainly comes from the English writing texts at the University stage. The corpus of low-end and high-end learners is scarce. Therefore, it is suggested that relevant researchers or institutions collect the writing corpus of these two groups extensively and establish the writing corpus of high-end and low-end English learners, so as to link up the writing corpus of each stage of English learning [8]. For the built corpus, researchers can supplement and improve the corpus by using the writing texts of recent large-scale examinations or students' daily exercises, so as to make the corpus more representative.

Secondly, after supplementing and improving the corpus, it is necessary to re-examine and revise the classification of corpus [9]. The first method is to consider learners' time to learn the target language and the degree of education; the second method is generally used by teachers to judge the level of students or their compositions and to divide them into different levels; the third method is to grade students according to their scores in standardized tests; and

the fourth method is to establish docking between corpus and well-known language competence scales, and to classify corpus according to the scales. Since China has not yet built a language proficiency scale that meets the characteristics of English learners in China, the fourth method can't be adopted for the time being. In the process of constructing the Writing Scale, the first three methods can be used to make a preliminary classification of the writing corpus.

Thirdly, after the preliminary classification of the corpus, the relevant corpus analysis tools and advanced statistical means can be used to analyze and compare the characteristics of different levels of text in the corpus from vocabulary, word frequency, collocation, cohesion, sentence length and so on. In this regard, we can refer to the research results of the EP project of the European Council [10].

Finally, according to the text features extracted from corpus analysis, the validity of the writing competence scale initially constructed by the "top-down" method can be validated, and on this basis, the scale can be supplemented, modified and improved to make the scale more scientific in classification and more detailed and specific in language competence description. Of course, after the completion of the writing competence scale, it can also be used to verify the classification of corpus. The scale and corpus are mutually validated and complementary, serving foreign language teaching and testing in China.

4. Prospects and Limitations of Corpus Application in Language Testing

4.1 Application prospects and trends

In the future language teaching, the use of corpus will become more and more frequent, and the use of corpus will be improved. In writing, automatic scoring based on corpus can be used to analyze the surface features of language. However, the automatic scoring system of spoken English needs further research and development, because it can only find out the typical characteristics of writing and speaking ability and the distinctive characteristics of different groups of candidates. By using corpus in the study of candidates' writing and speaking abilities, the candidates' answer strategies and new types of questions can be obtained. At the same time, in the investigation of language variants, the application of corpus in the field of modern language testing has broad prospects.

4.2 Limitations

In the field of language testing research, corpus can provide a lot of help, but corpus is only a kind of database, which can't include all the information that researchers need. Moreover, corpus also has some limitations in the application of information. Firstly, different reference points will appear when corpora are different, the parameters and purposes used in the design are different among corpora, so when using corpus, it is necessary to consider the representativeness and relevance of the samples contained in the corpus, otherwise misleading decisions and judgments will arise because of random use; secondly, the evidence provided by the corpus is not conclusive, if other information is used to verify and supplement the information involved in the corpus, professional theoretical knowledge should be used to reveal the corpus data. When using corpus for testing tasks, due to the limitations of test types and other requirements, it is necessary to consider whether other factors will affect its testing.

5. Conclusion

Corpus linguistics is becoming more and more mature, and corpus methods are also widely used in language research, language teaching research, language testing and other fields. Currently, language test developers and researchers have used corpus as an important tool for test development and scientific research. The use of corpus can make the test more real and accurate, and make the scoring scale more objective, reliable and generalized. Using corpus in writing test questions can make writing more efficient and time-saving, and also improve learners' learning efficiency. However, corpus is an innovative product with corresponding limitations, so we should be good at screening and perfecting in the process of using it. Several oral and written corpora have been established for English learners in China, but their application in foreign language testing is seldom discussed. Therefore, language testing researchers should recognize the potential of corpus in foreign language testing, and actively use this powerful tool to improve and perfect foreign language testing in China.

ACKNOWLEDGEMENTS

The study was supported by the Philosophy and Social Science Fund of Education Department of Jiangsu Province, Project No.2016SJD740035.

REFERENCES

[1] Kang Jun. A Corpus-based Study on the Legislation of Minority Languages in China. Journal of Hubei University of Arts and Sciences, 2019,40 (03): 36-40.

[2] Wang Xuan, Semati Metimin. A Corpus-based Data-Driven Approach to Language Learning. Cultural and Educational Materials, 2019 (07): 14-15+18.

[3] Niu Wenwenwen. Research on College English Language Teaching Guided by Corpus Linguistics. Modernization of Education, 2019, 6 (13): 124-126.

[4] Mei Qing. Cognitive Analysis of English Synonym Discrimination based on Corpus. Journal of Heilongjiang Institute of Education, 2018, 37 (10): 116-118.

[5] Su Heng. Application of Corpus in Language Testing Development and Design.Overseas English, 2018 (12): 216-217.

[6] Zou Shaoyan. The Application of Corpus in Language Testing: Review and Reflection. Journal of Ocean University of China (Social Science Edition), 2016 (06): 109-114.

[7] Cui Wei. On Corpus Application in Russian Testing. Northeast Asian Foreign Language Research, 2016, 4 (02): 60-64.

[8] Dong Yongyi. The Relationship between Corpus Linguistics and Language Testing. Language Construction, 2016 (14): 87-88.

[9] Zeng Yongqiang. Corpus-based Language Testing Research. English Teachers, 2015, 15 (22): 158.

[10] Zhang Wenxing. Application of Corpus in English Language Testing. Journal of Hunan Institute of Engineering (Social Science Edition), 2013, 23 (03): 51-53.