

Application of Corpus Analysis Tool in Investigating Listening Proficiency: Under the Perspective of TEM-8 Testing by RANGE

Yuting Chen^{1, a}, Shuang Liang^{1, b} and Ying Shen^{2, c *}

¹ Foreign Languages College, Guangxi University, China

² Foreign Languages College, Guangxi University, China;

Institute of Intelligent Systems, the University of Memphis, USA

^a 769663759@qq.com, ^b 745092514@qq.com, ^c sying@memphis.edu

* Corresponding author Ying Shen

Keywords: Corpus Analysis Tool; listening proficiency; TEM-8; RANGE

Abstract. With the development of the computer and the processing technology, various text analysis tools have been created. The purpose of this paper is to investigate how students can apply the software RANGE to the test-oriented education, helping them to improve their vocabulary levels in their listening tests. We used qualitative and quantitative approaches to analyze the listening exam items in Test for English Majors-Band 8 (TEM-8) from 1996 to 2010. The main findings are as follows: (1) The frequency of the tokens and types in the entire listening tests is found; (2) The common features in the vocabulary and the language of the listening tests become obvious; (3) The principles and features of the resources of the listening tests are known; and (4) Some of the strategies when taking TEM-8 listening tests should be exerted. This study can not only give some implications in TEM-8 listening testing but also shed some light on the development of students' English comprehensive ability.

Introduction

RANGE is corpus analysis software which can be used to analyze the breadth and the depth of the vocabulary in the texts. This software was designed by the linguists of Victoria University, Nation and Coxhead (2002), and compiled by Heatley. It has several vocabulary lists classified on the basis of different word grades, so this software can help researchers analyze the distributions of different word grades, frequency of word families according to given vocabulary lists. Its operation is convenient and easy. When plain text data is input to RANGE, it can analyze the data automatically in a few seconds. Anyone can download this free software through the website ([http://www.vuw.ac.nz/lals/staff/Paul Nation](http://www.vuw.ac.nz/lals/staff/Paul%20Nation)).

He (2001) used RANGE to retrieve the spelling error types and frequencies in the corpus which came from students.

Wen et al (2003) exploited RANGE to analyse the colloquial phenomena in students' compositions.

Li (2003) adopted RANGE to get the vocabulary statistics in the writings of students.

Bao and Wang (2005) conducted a case study and then drew a conclusion that the software RANGE can be used in the assessment of learners' productive vocabulary during the course of second language teaching. They found a way to apply RANGE to help the learners to improve their English writing abilities.

The objective of this study is to make a comparison study on the relationship between the source listening tests (TEM-8) and the authoritatively given word lists (West, 1953; Coxhead, 2002). By comparing a range or the distribution of the words in researched texts, researchers can be assumed to get the following information: a headword frequency figure, a word family frequency figure, and a frequency figure that a word occurs in tested texts. In addition, it is also considered that researchers

can create word lists based on the frequency that RANGE gives, and discover some common or the unique vocabularies in listening tests.

Research design

Subjects

In this study we take the listening parts of TEM-8 exams as the researching objects of RANGE. In other words, we used the listening tests of TEM-8 exam papers from the year 1996 to 2010 as the research source of this study.

Research tool

RANGE is used as a corpus tool, whose basic principle is to make a comparison between the researching object texts and the authoritative words lists. RANGE can be used to compare the vocabulary of up to 32 different texts at the same time. For each word in the texts, RANGE provides a range or distribution figure of a word, a headword frequency figure (a total time the actual headword type appears in all the texts), a family frequency figure (a total time the word and its family members occur in all the texts), and a frequency figure the word occurs in all the texts. RANGE can be used to find the coverage of a text based on certain word lists, create word lists in terms of frequency and range, and to discover shared and unique vocabularies in tested writings.

In RANGE there are three authority vocabulary lists, which were named as BASEWRD ONE, BASEWRD TWO and BASEWRD THREE. BASEWRD ONE contains 1000 word families (actually 999 word families), which are the most commonly used words. BASEWRD TWO contains 1000 word families (actually 987 word families), which are less common used than those in the BASEWORD ONE. These two basic vocabulary lists came from the top 2000 words in the “general English vocabulary list”, which was edited by West (1953). BASEWRD THREE collects the words which are not in BASEWRD ONE and BASEWRD TWO. Words in BASEWRD THREE include the most commonly used academic vocabularies that appear in various subjects or all kinds of teaching materials in the senior high schools and universities, whose number is reaching a total of 570 English word families. Actually, BASEWRD THREE is an academic vocabulary list edited by Coxhead (2002).

Results and discussions

First, it can be found that the abbreviations such like “s”, “t”, “ve”, “re”, “d” and “m” respectively rank 14, 43, 126, 152, 158, and 161 in the frequency results. These abbreviations often appear in the listening tests in forms of constitutes or a combination of letters, such as it’ s, he’ s, she’ s, I’ ve, you’ ve, we’ ve, you’ re, we’ re, don’ t, I’ d, I’ m. In spoken English, people often use contracted forms to take the place of the complete words or letters. Such is also the case in the listening texts.

Second, we found that one of the important parts of the speech knowledge is about dominant tone and weak phenomenon, and the speech knowledge also contains other parts: tone, pronunciation, stress, liaison, weak, incomplete explosion pronunciation rules, and so on. Only by mastering the phonetic knowledge can we be able to make a rapid, timely, and accurate response to the listening materials, especially when taking a difficult exam like TEM-8. Therefore we believe that students should practice their listening skills carefully in the parts of listening pronunciation, liaison, stress, weak, loss of explosion voice phenomenon, and other basic skills.

Third, the frequency of conjunctions is sequenced as follows: AND, OR, BUT, SO, WELL, FIRST, and NOW. These conjunctions play an important role of connection or transition of texts, and they are likely to be the key cohesive words to form the framework of the whole listening texts. With a good understanding of these conjunctions the students are assumed to quickly understand the internally logical relationship between the segments of the whole text structure, and the students can completely and accurately grasp the main points of the texts.

Fourth, the frequency of the interjections in TEM-8 listening tests is high to some extent. Some of the interjections with high frequencies are: NO (ranking the 17th), YES (ranking the 127th), YEAH (ranking the 223rd), OK (ranking the 232nd). Having many interjections is one of the language features in TEM-8 listening tests. These interjections are most likely to be used to reflect the speakers' attitudes and opinions. Therefore students should pay attention to the interjections when listening to the texts. However, in the actual listening training course, the same interjections will have various tones. If the tone is different the meaning will be different. This implies the speaker's "implication" in the interjections and the tone. As a result, students must listen carefully and distinguish the speaker's "implication" carefully. Only by accurately grasping the interjection of emotional intents of the speakers can students avoid making misunderstanding.

Summary

Based on the above analysis, we can see that the application of RANGE software in TEM-8 exams can be considered to be reliable. During the course of using RANGE, the students can apply it to different projects with different purposes. The data provided by RANGE is expected to be used to help students efficiently prepare TEM-8 tests. With the help of RANGE the students can judge the difficulties in different listening parts and thus be informed of the points of the difficulties. In addition, after analyzing the data provided by RANGE, the students will make sense about the features in the vocabulary use in the listening parts of TEM-8 exams. Moreover, the students can make comparisons and contrast between and against different texts in using words. Last, students can use RANGE to collect and edit the vocabulary lists according to their own needs since RANGE has a function of data analysis.

Acknowledgments

This research is financially supported by the following research foundations: 1). Core research foundation for postdocs at Guangxi University in 2010; 2). Innovative Research Foundation for National University Students in 2014 (NO141059349); 3). Distinguished University Teachers' Overseas Detachment Foundation in 2014-2015 by Guangxi Higher Education Bureau.

References

- [1] A. Coxhead. A new academic word list, *TESOL Quarterly* 2: pp. 213-238, 2002.
- [2] A. P. He. Analysis of student errors in spelling, *Foreign Language Teaching and Research* 3: pp. 199-205, 2001.
- [3] G. Bao and X. Wang. An application of RANGE in the evaluation of productive vocabulary, *Foreign Language Electronic Teaching* 8: pp. 54-58, 2005.
- [4] J. Li. A concise study on the construction and application of English abbreviations, *Journal of Qinhai Normal University* 4: pp.126-127, 2003.
- [5] M. West. *General Service List of English Words*. London: Longman, 1953.
- [6] Q.F. Wen. Oral tendency in English writings by Chinese university students, *Foreign Language Teaching and Research* 4, pp. 268-274, 2003.