

Discourse Variation between CET-6 and TOEFL Listening Tests and Implications for Teaching

Zhang Yan*

Jilin Engineering Normal University, Changchun, Jilin, 130062, China

**Corresponding author: zhangyanpan@126.com*

Keywords: CET-6; TOEFL; Listening tests; Discourse analysis; Lexical structure

Abstract: This study takes CET-6 and TOEFL as the research objects. Text Inspector and Lexical Tutor are used to conduct discourse analysis on the listening parts of the two tests, and the focus is on the variation in lexical structure. The results show that there are obvious differences between CET-6 and TOEFL listening in lexical features, subjects, modalities, and pause fillers. Based on the results, this study discusses the linguistic authenticity of the teaching and testing materials, and provides suggestions for the authors, teachers, and learners.

1. Introduction

Although there are various differences between CET-6 and TOEFL listening in terms of content, design, method, and administration, there are similarities in their testing philosophies. The goal of the listening part of CET-6 is to test the examinees' communicative awareness and communicative ability in study, work, and life. The listening section of TOEFL is designed to test the examinees' language and communication skills by mimicking a real language environment at the universities in North America. They both attach great importance to the communicative function of language. Although the listening materials of the past exam papers of the two tests are highly frequently used by learners, few studies have done discourse analysis on them, and few studies have paid attention to the differences in lexical structure between CET-6 and TOEFL listening. To this end, this study aims to answer the following research questions:

- (1) What is the discourse variation between the listening part of CET-6 and TOEFL in terms of lexical structure?
- (2) What are the implications for teaching?

2. Literature Review

The comparative studies of different English tests mainly focus on two aspects. Some studies compared CET, IELTS, and TOEFL at a macro level. They probe into the similarities and differences of these tests in content, design, methodology, and administration.^[1] For example, Tang compared CET with IELTS in terms of the task types and abilities to be measured.^[2] More researchers focus on one of the four aspects of listening, speaking, reading, and writing. For instance, Wang and Gu conducted a comparative study on the difficulty levels of the reading parts

of CET-6, IELTS, and TOEFL.^[3] Liang conducted a comparative study on the backwash effect of the speaking parts of the above-mentioned tests.^[4] Hong conducted a similar study, focusing on the writing part.^[5]

Although there are studies comparing various English tests, most of them concentrated on macro level comparison. If language is regarded as a discourse, vocabulary must receive the same attention as other forms of language.^[6] This study analyzes the differences between CET-6 and TOEFL listening tests, specifically long conversation listening, from the perspective of lexical structure, with the hope of bring implications to teaching.

3. Methods

This study built two corpora for CET-6 and TOEFL listening materials respectively. Using the retrieval function of the corpus and online corpus analysis tools, it compared the lexical structure of CET-6 and TOEFL long conversation listening on four levels: lexical features, subjects, modal verbs, and pause fillers. The specific research methods are as follows.

3.1 Data Collection

Since the length of TOEFL long conversation listening is longer than that of CET-6, and to ensure that the two corpora are equal in size, this study selected 14 CET-6 long conversation listening materials from December 2017 to December 2020 and 10 TOEFL long conversation listening materials from TPO 62 to 66 randomly. All the data are proofread with reference to the audio to ensure the accuracy of data.

3.2 Research Instruments

In this study, Text Inspector, an online text and lexical analysis tool, and Compleat Lexical Tutor, a free online English corpus, were used to analyze the lexical features of the data. Subjects, modalities, and pause fillers were searched successively by AntConc corpus retrieval software, and Log-Likelihood (LL) was calculated. A LL value higher than 3.84 indicates a significant difference. The higher the value, the more obvious the difference.

4. Results

The results show that there are obvious differences between CET-6 and TOEFL listening in terms of lexical features, subjects, modalities, and pause fillers.

4.1 Differences in Lexical Structure

The study adopted Text Inspector and Lexical Tutor to analyze the lexical diversity, type-token ratio, lexical density, and ratio of academic words of the data in the two corpora respectively, and the results are shown in Table 1.

Table 1: Comparison on the Lexical Feature between CET-6 and TOEFL Listening

	Lexical Variance	Ratio of Academic Words	Type-token Ratio	Lexical Density
CET-6	49	4.46%	26%	72.22
TOEFL	47	2.97%	21%	65.68

Lexical variance is used to measure the number of different words in a corpus, and is an

important indicator of lexical complexity and text difficulty. The type-token ratio can also reflect the lexical variance to some extent. The larger the value, the larger the number of different words. The smaller the value, the higher the repetition rate of words. As shown in Table 1, the lexical variance and type-token ratio in CET-6 are higher than those in TOEFL, indicating that CET-6 has more rich and diverse vocabulary, while TOEFL has less lexical diversity and slightly higher repetition rate of words. While having dialogues, due to the pressure caused by short-term memory, speakers are accustomed to calling the most familiar expressions in memory. Some fixed and limited number of lexical chunks are used repeatedly in communication, which are typical characters of spoken conversation.^[7] The lexical density of CET-6 is also higher than that of TOEFL. It is generally believed that the higher the lexical density, the more formal the discourse. In addition, the ratio of academic words in CET-6 is significantly higher than that of TOEFL. Tian believed that academic words enjoy a high proportion in academic texts and is generally used for specific purposes in English teaching.^[8] Based on the above analysis, the long conversations in listening tests should be colloquial in style. CET-6 long conversations are richer in terms of word choices, with a higher proportion of academic words, and is more like written language. On the other hand, the vocabulary in TOEFL long conversations has little lexical diversity. Besides, basic daily words are used, which is a typical character of spoken English.

4.2 Differences in Subjects

In this study, AntConc corpus was employed to analyze the frequency and ranking of subject occurrence in the two corpora, and the LL values are calculate. It is worth noting that the study mainly analyzes the subject from the perspective of systemic-functional grammar, so the unit of the subject is a clause rather than a whole sentence. In some contexts, the subjects of some sentences are specific nouns, verb phrases, etc., which are not of analytical value. This study focuses on the comparison of personal pronouns as subjects, and the specific results are shown in Table 2.

Table 2: Comparison of Subjects between CET-6 and TOEFL

Subject(Personal Pronoun)	CET-6		TOEFL		LL
	Ranking	Frequency	Ranking	Frequency	
I	6	105	1	197	17.39****
we	18	40	32	27	2.52
you	4	102	3	147	0.77
she	83	10	N	0	15.28****
he	148	5	71	12	2.09
it	14	64	12	66	0.36
they	29	28	65	14	6.86**

Personal pronouns are important indicators to realize specific communicative functions. Because of their high occurrence, personal pronouns have become factors that cannot be ignored in discourse analysis. It can be seen from the above table that in TOEFL long conversation listening, the first person “I” appears most frequently, and ranks the first among all the subjects. In the CET-6 corpus, “I” only ranks sixth. The LL value is 17.39, indicating that there is a significant difference between the two tests in the use of “I”. Another significant difference is the use of the third person “she” and “they”. The ranking and frequency of “she” and “they” in the CET-6 corpus are higher than those in TOEFL, and the LL values also show significant differences. Therefore, it can be concluded that CET-6 listening uses the third person more, and the dialogue is mostly about other people or things, while TOEFL dialogues use the first person more, and the participants are mostly both sides of the dialogue. Biber used a corpus to conduct quantitative research on 481 texts, showing the features of

different types of discourse by drawing quadrant, in which the position of dialogue-type discourse is close to the highest value of “engagement”, indicating that one of the obvious characteristics of dialogue-type discourse is the high engagement of the speakers.^[9] From this point of view, compared with CET-6, TOEFL listening has more characteristics of dialogue.

4.3 Differences in Modalities

Leech believed that dialogues emphasize interpersonal interaction, in which linguistic expressions representing emotions and attitudes are bound to appear repeatedly.^[10] Halliday believed that every modality has its own semantic value, and divided them into three categories of high, median, and low according to their semantic values.^[11] Modality of high value include “must”, “be required to”, “have to”, “mustn’t”, “can’t”, etc. Modality of median value include “will”, “would”, “should”, “won’t”, “wouldn’t”, “shouldn’t”, etc. Modality of low value include “may”, “can”, “could”, “needn’t”, “don’t have to”, etc. This study analyzes the frequency of modalities of each category in CET-6 and TOEFL listening, and the results are shown in the table below.

Table 3: Comparison of Modalities between CET-6 and TOEFL

	Modality of High Value	Modality of Median Value	Modality of Low Value
CET-6	20	23	35
TOEFL	19	26	56
LL	0.06	0.11	4.3*

One of the typical features of natural verbal discourse is the use of modalities to express interactive relations. The analysis of modalities will help to understand the speaker’s attitude towards the listener and the thing being talked about. It can also indicate the social distance between the interlocutors and the degree of formality of the conversation. As shown in Table 3, there is no significant difference between CET-6 and TOEFL in the use of modalities of high value. The main difference is reflected in the frequency of the use of modalities of low value. The appearance of modalities of low value indicates that the speaker has a lower degree of responsibility and expectation for the conveyed information, and the discourse is remain to be discussed, leaving more room for the listener. From this point, compared with CET-6, the expressions in TOEFL listening is more euphemistic, and more indirect forms of interpersonal interaction are used. This vagueness and indirectness are the conversation habits in real life.

4.4 Differences in Pause Fillers

Pause Filler is a type of discourse marker. Wang and Zhu divided discourse markers into logical connection markers and pause fillers. Pause fillers are further divided into fillers with specific meanings (e.g. I mean...) and fillers without clear meanings (e.g., well, oh).^[12] This study focuses on the use of pause fillers in CET-6 and TOEFL corpora.

Table 4: Comparison of Pause Fillers between CET-6 and TOEFL

	Fillers with specific meanings	Fillers without clear meanings
CET-6	47	23
TOEFL	109	133
LL	17.59****	71.48****

During a conversation, thinking and output happen almost simultaneously, and the speaker needs to search for an expression quickly in a short time, which will inevitably show hesitation or stalling. Therefore, the high occurrence of pause fillers in spoken conversations is a natural phenomenon. The results in the table above show that in CET-6 corpus, pause fillers appear 70 times, while in the

TOEFL corpus, they appear 242 times, three times as many as in CET-6. Thus, the most significant difference in lexical structure between CET-6 and TOEFL listening is the use of pause fillers. Pause fillers frequently appear in listening and speaking context, which makes the conversation more natural. However, CET-6 listening pays more attention to the transmission of information, ignoring the imitation of the scene of a real dialogue. This insufficient use of pause fillers affects the naturalness of dialogue to a certain extent, so that the performance of communicative function is compromised.

5. Discussion and Implications for Teaching

Based on the above results, it can be concluded that there is not much lexical diversity in TOEFL listening, the content is mostly centered on the two sides of the conversation, the pause fillers are frequently used, and the discourse is more close to the dialogue in real scene. On the other hand, the CET-6 dialogues pay more attention to the output of information, using rich and varied words. The academic vocabulary ratio is high, and pause fillers are less used. The conversations are formal in style and slightly rigid in content. It has a prominent features of “human-made”, and lack of authenticity in discourse.

According to Timlinson, English testing materials are highly used among Chinese learners, so they can be regarded as teaching materials.^[13] The authenticity of teaching materials has always been a hot topic in the field of English education. The authenticity concerned in this paper does not refer to whether the discourse is taken from real data, but whether the discourse has real oral characteristics. Mary Underwood, Assistant Dean of Ealing College of Higher Education in London, mentioned in a lecture that the discourse used in listening training should have natural rhythm, intonation, pronunciation, interjection, and repetition, and the language style should be colloquial, starting and ending naturally. However, some listening materials are unnatural in rhythm and intonation. There is no overlapping or interjection in the dialogue, and the language style is written-like.^[14] Hymes proposed that one of the purposes of language teaching is to equip students with communicative competence, and the fundamental goal of listening learning is to enable learners to demonstrate communicative competence in real life scenarios.^[15] From this point of view, the selection of some listening materials and past exam papers runs counter to the principle of authenticity.

Based on the above discussion, this study suggests that the compilers of listening textbooks and testing materials should pay attention to the interpersonal functions of the discourse, present the oral features of dialogues as much as possible on the basis of ensuring the transmission of language knowledge and information, and improve learners' awareness of the oral features of the discourse, so as to acquire authentic and natural expressions. For teachers, one of the problems they have to face is to carry out listening teaching according to the established teaching syllabus and textbooks, and teachers need to make appropriate analysis, adjustment, and modification of the possible materials which are out of the characteristics of real dialogue. Discourse analysis can play an important role in the process of researching and evaluating the authenticity of the materials, and tools such as corpus can also provide references for the revision and adjustment of materials. In addition, teachers can not only provide learners with more real and natural listening input, but also guide them to compare the teaching materials or listening materials in the past exam papers with real oral dialogues to enhance their awareness of the differences.

6. Conclusion

Taking CET-6 and IELTS listening as the research object, this study uses corpus and discourse analysis tools to explore the differences in lexical structure between the two tests. From the

perspective of cultivating learners' communicative competence, this study discusses the textual authenticity of teaching and testing materials, and advocates the compilers of teaching and testing materials, teachers, and learners to raise their awareness of discourse differences. Modern tools such as discourse analysis tools and corpora should be used to make up for the lack of authenticity in teaching materials and exam papers, so as to truly improve learners' ability to communicate in real contexts.

References

- [1] Ma J. (2006). *A comparative study of TOEFL, IELTS, and CET[R]*. A master's degree thesis, central China normal university.
- [2] Tang J. (2016). *Comparison of CET-4, CET-6, IELTS and TOEFL [J]*. *English Abroad*, (24), 72-73.
- [3] Wang P., & Gu X. D. (2020). *A Comparative Study on the difficulty of reading texts in CET-6, IELTS and TOEFL—A method based on data mining [J]*. *Foreign Languages and Translation*, (04), 11-16.
- [4] Liang Y. S. (2017). *A Comparative Washback Study of CET-SET6, IELTS and TOEFL Speaking Tests: Evidence from Test-takers' Perspectives [R]*. A master's degree thesis, Chongqing University.
- [5] Hong Y. (2017). *A Comparative Washback Study of CET-SET6, IELTS and TOEFL Writing Tests: Evidence from Test-takers' Perspectives [R]*. A master's degree thesis, Chongqing University.
- [6] Mccarthy M., & Carter R. (1994) *Language as discourse perspective for language teaching [M]*. New York: Longman Publishing.
- [7] Zhang Y. L., & He A. P. (2009). *Oral Inquiry into English textbook dialogue [J]*. *Foreign Language Teaching Theory and Practice*, (02), 61-67.
- [8] Tian M. (2020). *A study on the effectiveness of academic English vocabulary in the writing of non-English majors [J]*. *Journal of Liaoning Economic Management Executive College*, (02): 137-139.
- [9] Biber D. (1991). *Variation across speech and writing [M]*. Cambridge: Cambridge University Press.
- [10] Leech J. (2000). *Grammar of spoken English [J]*. *Language Learning*, (5): 675-724.
- [11] Halliday M. (2004). *Introduction to functional grammar [M]*. London: Arnold.
- [12] Wang L. F., & Zhu W. H. (2005). *A study on the use of discourse markers in Chinese students' oral English [J]*. *Foreign Language Studies*, (03), 40-44, 48.
- [13] Tomlinson B. (2001). *Introduction: principles and procedures of materials development [M]*. Cambridge: Cambridge University Press.
- [14] Zhang Z. X. (1985). *Real and unreal teaching materials for listening training [J]*. *Foreign Language Teaching*, (02), 25.
- [15] Hymes D. (1967). *Models of the interaction of language and social setting [J]*. *Journal of Social Issues*, (01): 8-38.