

Corpus-Based Study on Dictation in TEM4 Preparatory Materials

ZHANG Yang¹, YANG Chen²

¹Doctoral candidate, Faculty of Humanities, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR

²Sino-foreign joint programme manager, International Education College, Henan Polytechnic University, Jiaozuo, P. R. China

Correspondence: ZHANG Yang, Doctoral candidate, Faculty of Humanities, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR.

Received: March 13, 2019

Accepted: March 26, 2019

Online Published: April 9, 2019

doi:10.5430/wjel.v9n2p1

URL: <https://doi.org/10.5430/wjel.v9n2p1>

Abstract

Corpora has been widely used in language teaching and learning nowadays. The use of corpora facilitates teachers for effective language teaching, meanwhile it provides valuable materials for the language learners to learn the language. The current study focuses on the self-designed corpus in dictation part in TEM4 and aims to explore if there any difference exists in the three sets of materials (TEM4 dictation, TEM4 mock and listening textbook dictation). The author finds that TEM4 dictation enjoys higher quality compared with the other two types of the material. In addition, the listening textbook dictation covers most topics that exist in TEM4 dictation. The mock test dictation fails to cover as much relevant information as possible. Such findings will be helpful for the test-takers to prepare the TEM4 dictation.

Keywords: TEM4 dictation, self-designed corpus, language teaching, language learning

1. Introduction

Since the implementation of the TEM4 syllabus in 1993, dictation, as one of the test tasks in the listening section, plays an integrative and indispensable role in TEM4. It aims to test students' familiarity with vocabulary and sentence patterns required by the TEM4 syllabus (Zou, 1996). In the past two decades, a proliferation of research has been conducted on TEM4. Most of them concentrate on writing and reading parts in the test. Little has been devoted to the listening section, especially for the dictation part. Based on the three preparatory dictation materials (TEM4 dictation, TEM4 mock dictation and listening textbook dictation) the current study first builds self-designed corpus with the aid of the corpus software. This study aims to explore the differences of TTR and the frequency of nouns in three sets of materials. It then concludes the findings and tentatively offers suggestions regarding TME4 dictation in hope of providing useful information for teachers and helping learners better prepare for the test.

2. Literature Review

Corpora enjoy higher popularities in discourse analysis (Piper, 2000), according to Sinclair (Sinclair 1991:171). A corpus is "a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language." The analysis of the text based on the corpus provides opportunities for researchers to investigate the frequently used patterns or rare, occurring language phenomenon in the large-scale context. Many corpus software helps improve efficiency and save time by providing functions such as words frequency lists, keywords lists, keyness, etc.

It is admitted that the word frequency list could be used not only in different varieties of English but also in different contexts. In terms of the varieties of English, James et al. (1994) published a book regarding the frequency words list in computer fields. Based on the MICASE, the professional team in Michigan produced the frequency wordlists for academic spoken English in America. In language teaching context, the use of words frequency in the educational setting has been discussed by many researchers. Fires and Traver (1950) built the wild range of vocabularies to investigate the possible applications in educational settings. Recently, one of the representatives is Schmitt and McCarthy (1997) who connected closely language teaching with the frequency words lists. In addition to the frequency lists in TEM4 dictation, this study also uses the comparison of the data that shares a similarity in size (Granger, 1998) in the exploration of the words variation in the text.

The use of keyness in corpus analysis to find the most frequently used words has been adopted by many researchers. Tribble (2000) made a comparison between fiction corpus and generally spoken corpus, which helps him find possible explanations of the language feature in this specific context. Johnson et al. (2003) adopted keyness to investigate the use of keywords in newspapers, which foregrounds the characteristics contained in the particular context. Due to the corpora capacity and workload, a few insightful researches have been done in this area. Based on the previous study, this research aims to explore the use of keywords in the listening context in TEM4.

3. Research Questions

1. What are the differences of TTR in preparatory dictation materials for TEM4, more specifically:
 - (1) What are TTRs of dictation in three materials?
 - (2) What are TTRs of dictation in four different periods of TEM4?
2. What types of nouns are frequently used in Dictation of preparatory materials for TEM4, more specifically:
 - (1) What types of nouns are frequently used in TEM4 dictation?
 - (2) What types of nouns are frequently used in TEM4 mock test dictation?
 - (3) What types of nouns are frequently used in listening textbook dictation?
 - (4) What are relationships among the three preparatory dictation materials regarding the frequently used nouns?

4. Data & Methodology

4.1 Instrument

Antconc (Anthony, 2018) is employed in the current research in order to investigate the differences of TTR and words frequency in Dictation of preparatory materials for TEM4. The analysis of TTR and words frequency primarily relies on Antconc for the sake of improved work efficiency. Also, a comparison is made on the differences of three materials.

4.2 Introduction to Data

The current study is conducted based on the collected dictation materials, namely, TEM4 dictation from 1993 to 2018, TEM4 mock test dictation from 2005-2018 and listening textbook dictation (Listen to This 2). The implementation of new test syllabus for TEM4 in 2004 proposed new test requirements and orientation. In this sense, the selection of dictation materials for mock test in this study is a reflection on the common practice in the past 13 years. With a view to the differences and qualities of dictation materials available in current teaching environment, the study chooses the materials compiled by New Oriental English which enjoys a higher reputation in Mainland China. After constant discussions with colleagues, dictation materials are selected randomly on the New Oriental English TEM4 self-learning website. Listening textbook called "Listen to This 2" is selected as reference material in this research. As one of the traditional listening textbooks in line with the test requirements for listening parts in TEM4, this textbook is designed primarily for those with intermediate-level language proficiency.

4.3 Data Collection

Altogether 110 texts are collected in the current research, namely TEM4 dictation (25), TEM4 mock test dictation (25), and listening textbook dictation (25). The three types of collected data are respectively documented in text format to facilitate the use of corpus tools.

Table 1. Data Collection

Text	Number
TEM4 dictation	25
TEM4 mock test dictation	25
Listening textbook dictation	25

5. Findings & Discussion

5.1 TTR of Dictation in Three Types of Collected Data

TTR of dictation in three types of collected data is analyzed with the aid of Antconc. The table below displays words types, tokens, and T/T ratio in each type of data.

Technically and theoretically, the three types of data are interrelated with each other since all of the materials are used by English majors in their preparation for TEM 4. Moreover, vocabularies selected in these three types of

materials should share similarities in the same level so as to be in accordance with the requirements in test syllabus. Therefore, they share similarities in some points such as vocabulary difficulties and vocabulary range. Such similar patterns could be further represented in types, tokens, and ratio.

Table 2-a. TTR of Dictation in three materials

Contents	A1	A2	A3
Word Types	1159	1000	1400
Word Tokens	3811	4600	4200
T/T Ratio	30.4	21.7	33.3

(Note: A1=TEM4 dictation, A2= TEM4 mock test dictation, A3= Listening textbook dictation)

Types

Form table2-a, it is found that the word types in each kind of material are A1 (1159), A2 (1000), A3(1400). A1 and A2 are almost the same regarding the word types, and the number of word types in A3 is the largest among the overall collected materials, which increases to 241 words types in comparison with A1. Such differences between A1 and A3 may reflect that listening textbooks provide such little vocabulary for dictation exercises practice that students' familiarities with vocabularies will be enhanced and thus feel less stressful in TEM4.

Tokens

Table2-a also illustrates word tokens of overall collected materials, namely: A1 (3811), A2 (4600), A3(4200); The average words for each dictation passage in each type of collected material would be: A1 (3811/25=152 words), A2 (4600/25=184 words), A3 (4200/25=168 words). The calculated average numbers indicate that length of passage in both A2 and A3 are relatively higher in comparison with that in A1, which was beyond the words' requirement for dictation in the TEM4 syllabus (approximately 150). It seems that daily practice of dictations provides more information and knowledge than TEM4 tests. Such practice helps improve student's adaptability to TEM4 dictation and get familiar with the real test.

TT/Ratio

TT/Ratio reflects the difficulties and the intensity of the selected words in the overall materials. As is shown in table 2-a, TT/R of each collected data is 30.4, 21.7, 33.3. A3 is on the top of the three ratio which reveals a fact that TEM4 dictation is the most difficult and the listening textbooks are almost the same as TEM4 dictation in terms of the vocabulary difficulties. In addition, it reflects the design of the textbook which is more valid and conforms to the syllabus for TEM4. In comparison with the other two materials, the statistics show that the quality of TEM4 mock test dictation is relatively lower.

Table 2-b. TTR of Dictation in TEM4

Contents	B1	B2	B3	B4
Word Types	301	517	634	163
Word Tokens	618	1235	1707	278
T/T Ratio	48.7	41.9	37.1	58.6

(Note: B1=TEM4 dictation 93-96, B2= TEM4 dictation 97-04, B3= TEM4 dictation 05-15, B4= TEM4 dictation 16-18)

Table 2-b provides more detailed information of TEM4 test in terms of types, token and TTR. In the past two decades, TEM4 has witnessed three big syllabus reforms in the respective year of 1996, 2004 and 2015. Therefore, it is necessary to investigate specific aspects engendered by the three types of change. Considering different numbers of dictation passages in every syllabus reform period, the author only focuses on tokens and ratio as the two major sectors.

Tokens

The table shows that tokens in each period are B1(618), B2(1235), B3 (1707), B4(278), to calculate the average words for each passage, we get the following data: B1 (618/4=155 words),B2(1235/8=154 words), B3(1707/11=155 words), B4(278/39=92words). Based on the calculated average number, we could conclude that the length of each passage almost remains the same from 1993 to 2015. However, the length of the passage sharply declines from the year 2016. Such findings echo the implementation of the new syllabus for TEM4 in dictation (TEM Test Office, 2015): the dictation passage length should around 80-90 words.

T/T Ratio

The statistics indicate that during the past two decades, selected words in TEM4 are experiencing fluctuant trend. TTR in the fourth syllabus reform is the highest (58.6), and B1 comes to the next with TTR of 48.7. There is a

declining trend between 1997 and 2004, with TTR of 41.9 and 37.1 respectively. Although word length has decreased, the requirements for word difficulties and the words intensities become stricter than before. In other words, students should make more efforts in preparation for dictations if they would like to earn satisfactory grades.

5.2 Word Class Frequency

This study mainly focuses on the frequently-used nouns in each selected material, which may reflect the central theme in TEM4. According to TEM4 Syllabus (TEM Test Office, 2015), dictation contains popular topics such as people, world, and life. Nouns which are frequently used fall into several categories concerning the TEM4 syllabus. The tables below illustrate the detailed data information:

Table 3-a. Frequency of nouns in TEM4 dictation

Frequency	Nouns
37	people
17	education,
15	pollution, tourism,
9	English, friends, language,
8	children, examination. festival,
7	convention, finance, money,
6	election, music, trade
5	sports, university,
4	knowledge

The above table 3-a reveals that nouns of various kinds could be found in TEM4 dictation. Also, it indicates that different topics are involved in TEM4 dictation. Words like “people” “education” “pollution” and “tourism” have been mentioned many times with the frequency of 37, 17, 15,15 respectively. Such high frequencies in dictation foreground the importance of the topic closely related with students’ life. It is noted that the frequently-used words could also be regarded as main topic categories in TEM4 dictation. Each main category encompasses several keywords. Thus, we further make classification of frequently-used nouns according to themes. Firstly, when it comes to the topic “people,” it contains several subcategories such as people (37), friends (9), children (8). Next comes the topic “education”: education (17), examination (8), university (5), knowledge (4). Moreover, the “Culture and social life” topic covers the following nouns: tourism (15), festival (8), convention (7), music (6), sports (5). Under the topic of “country and language” contain language (9), English (9). The “economic and politics” topic comprises finance (7), money (7), election (6), trade (6). Lastly, “environmental problems” topic encompasses pollution (15). The collected data indicates that a wide variety of themes are selected in TEM4 dictation, which correlates with students’ life

Table 3-b. Frequency of nouns in TEM4 mock test dictation

Frequency	Nouns
25	education
17	people
11	University
9	noise, pollution
5	examination

As is mentioned in the above table 3-b, the top three of the frequently-used nouns are education (25), people (17), University (11) respectively. The next group comes to the following two nouns which share similarities in frequency, that is noise (9), pollution (9). Nouns with a frequency of 5 encompasses examination (5). Revealed by the statistics, the selection of themes in TEM4 mock test dictation is relatively limited in comparison with TEM 4 dictation. Most of the selected topics concentrate on “education” and “environmental” problems. In this case, the topics included in the mock test remains to be varied. Also, there remains an inadequacy in material useful for students in their preparation for the TEM4 test.

Table 3-c. Frequency of nouns in listening textbook dictation

Frequency	Nouns
24	campus, people, university
18	festival, sports, travel
15	pollution, media
14	noise
9	children
7	adults, campus, examination
6	students, knowledge
5	book, computer, comedy, tourism
4	democratic, election, finance, money, trade

Table 3-c illustrates the frequently-used nouns in listening textbook dictation. The top one nouns regarding frequency are people (24), campus (24), university (24), which is in line with the high frequency in TEM4 dictation. In other words, topics selected in listening textbook dictation are almost the same as those selected in TEM4 dictation and thus enjoys high quality and validity. Similarly, the classification of the additional nouns is based on the following main themes including people, education, culture and social life, environmental problems, economy and politics, and technology. Under the category “people”, there are people (24), children (9), adults (7) and students (6). Themes such as campus (24), university (24), campus (7), examination (7), knowledge (6), and book (5) are involved in “education” categories. Thirdly, “culture and social life” category encompasses the following themes, such as festival (18), sports (18), travel (18), media (15), comedy (5), tourism (5). For the topic “environmental problems” are pollution (15), noisy (14). In addition, “economy and politics” topic contains: democratic, election, finance, money and trade, which are all has the same frequency (4). Lastly, computer (5) are involved in “technology” category. The abundant themes in listening textbook dictation orient students themselves to the themes possibly encountered in real test and thus help them achieve an ideal score in the actual exam.

Table 3-d. Comparison of themes in three collected materials

Themes	Materials	TEM4 Dictation	TEM4 Mock Test Dictation	Listening Textbook Dictation
	People	people (37) friends (9) children (8)	people (17)	people (24) children (9) adults (7) students (6)
	Education	education (17) examination (8) university (5) knowledge (4)	education (25) University (11) examination (5)	as campus (24) university (24) campus (7) examination (7) knowledge (6) book (5)
	Culture & social life	tourism (15) festival (8) convention (7) music (6) sports (5).		festival (18) travel (18) sports (18) media (15) comedy (5) tourism (5)
	Environmental problems	Pollution (15).	pollution (9) noise (9)	pollution (15) noisy (14)
	Economy and politics	finance (7) money (7) election (6) trade (6)		democratic (4) election (4) finance (4) money (4) trade (4)
	Country and language	English (9) language (9)		
	Technology			computer (5)

Table 3-d displays the different themes involved in the collected materials. It is found that both TEM4 and listening textbook dictation cover every theme mentioned, while TEM4 mock test dictation only contains two themes: people and life. There exists specific information under each theme. And the number of the selected themes varies among different materials. Given the number of selected topics under each theme in the three materials, listening textbook dictation rank first, which encompasses as much information as possible to help students prepare for the dictation test. Next comes the TEM4 dictation where many themes are closely related to personal life. The last is mock test dictation which involves a relatively limited number of themes. Despite the fact the themes selected in mock test dictation are by one in TEM4 dictation, it fails to guide students to prepare adequately for TEM4 dictation. Also, in comparison with listening textbook dictation, the quality of TEM4 mock test dictation remains to be improved.

According to Buck (2001), a better understanding of the text background contributes to more accurate information. Therefore, topics in TEM4 dictation helps guide test-takers in their preparation for the test and thus obtain effective grades. As is suggested by Zou & Kong (2013), students are highly recommended to get familiar with various kinds of topics and make full use of the materials at hands in their preparation for the test. Also, the repetition of the noun words in the dictation directly influences students’ vocabulary learning (Chen, 2011), which is a necessary step for language acquisition (Ellis, 2001). In other words, the higher frequency of nouns in dictation contributes to effective language learning among students.

6. Conclusions

Regarding word types discussed above, listening textbook dictation covers more vocabulary in comparison with TEM4 dictation. Vocabulary in mock test dictation is relatively less compared with the other two materials. Such findings reflect the quality and validity of the content provided for students in their preparation for TEM4. For word tokens, both mock test dictation and textbook dictation comprise more words in each passage than what is required in the TEM4 syllabus. This helps students' familiarities with dictation content and contributes to active learning. Given TTR in the three materials, TEM4 dictation ranks first, followed by listening textbook dictation. Vocabulary in TEM4 is the most difficult in comparison with the other two materials. While textbook dictation is the same as TEM4 dictation regarding the vocabulary difficulty, the findings is in accordance with what is claimed by Oakland and Lance (2004). The higher the TTR is, the more difficult the text would be. Findings regarding TTR in the three materials also indicates that mock test dictation fails to conform to the test syllabus with a view to the vocabulary selection.

In the investigation of TEM4 dictation in different four periods, words length in TEM4 dictation are found to be increasingly reducing, especially after the brand-new reforms of TEM4 in 2015 when the selected texts are becoming much more difficult than before. Such changes in TEM4 dictation indicate that burden imposed on students is so heavier to push them towards full preparation for dictation.

For the frequency of nouns existed in three materials, listening textbook dictation comprehensively covers the most nouns in comparison with the other two materials. The frequency of nouns reflects the different kinds of topics or themes covered in dictation. Technically, the earlier students get involved in various kinds of topics in the preparatory period, the more familiar they may get with the real test (Shen, 2015). Abundant topics encompassed in textbooks guide students to make full preparation for TEM4 dictation while the mock test dictation fails to cover as much relevant information as possible, which fails to remind students of the trend embedded by the real test.

7. Limitation & Suggestions

The current study has several limitations listed as follows:

Firstly, only small numbers of materials are employed due to the limited time. Findings based on the analysis of the collected data may not be persuasive and convincing.

Secondly, this study only concentrates on the two points TTR and frequency of nouns while there are still many aspects in dictation of TEM4. A more comprehensive study on dictation needs to be conducted.

Thirdly, regarding the validity of the data, only one type of mock test is used in the study. Different versions of mock tests and the selection of materials are restricted by personal teaching experience and students' feedbacks. Future research may select as many versions of the mock test as possible to ensure the quality and validity.

According to the findings, suggestions are made as followings:

Firstly, students should pay closer attention to TEM4 dictation and listening textbook dictation, which could be used as guidance books in the preparation for the TEM4 dictation.

Moreover, teachers are highly advised to prepare as many valuable materials as possible for students to help them get familiar with the test. In the selection of such materials, a variety of vocabulary and themes of various kinds should be considered.

Lastly, a variety of themes should be considered in mock test dictation with the reference to the frequent topics and words in real tests.

References

- Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>
- Buck, G. (2001). 2001: Assessing listening. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732959>
- Chen, X. X. (2011). Case Study of Vocabulary Learning for English Major. *Foreign Language Teaching Forum*, 2, 9-14.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188. <https://doi.org/10.1017/S0272263102002024>

- Fries, C. C., & Traver, A. A. (1940). English word lists; a study of their adaptability for instruction.
- Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. *Learner English on Computer*, 119-131.
- James, G., Davison, R., Cheung, A. H. Y., & Deerwester, S. (1994). *English in computer science: A corpus-based lexical analysis*. Hong Kong University of Science and Technology.
- Johnson, S., Culpeper, J., & Suhr, S. (2003). From politically correct councillors' to Blairite nonsense': discourses of political correctness' in three British newspapers. *Discourse & Society*, 14(1), 29-47. <https://doi.org/10.1177/0957926503014001928>
- Oakland, T., & Lane, H. B. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing*, 4(3), 239-252. https://doi.org/10.1207/s15327574ijt0403_3
- Piper, A. (2000). Some have credit cards and others have giro cheques: individuals' and people as lifelong learners in late modernity. *Discourse & Society*, 11(4), 515-542. <https://doi.org/10.1177/0957926500011004004>
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge University Press.
- Shen, Y. (2015). A Case Study of Vocabulary in TEM Batteries. *Education Teaching Forum*, 5, 140-150.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- TEM (Test for English Majors) Test Office. (2015). TEM4 test report for 2015.
- Tribble, C. (2000). Genres, Keywords, Teaching: towards a Pedagogic Account of the Language of Project Proposals. *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt am Main: Peter Lang Pub, 76-90.
- Zou, S. (1996). Analysis of TEM4 Dictation. *Foreign Language World*, 1, 55-61
- Zou, S., & Kong, J. F. (2013). Analysis of TEM4 dictation in. *Foreign language Teaching and Testing*, 4, 1-6.