

A Preliminary Analysis of Text Coverage and Extensive Reading of Chinese ESL Learners-- Range Analysis of *The Little Prince* and *Pride and Prejudice*

Jiayu Zhuo*

Language Academy, University of Malaysia Technology, Malaysia

ARTICLE INFO

Article history

Received: 3 August 2021

Revised: 10 August 2021

Accepted: 20 October 2021

Published Online: 30 October 2021

Keywords:

Text coverage

Extensive reading

Vocabulary teaching

ESL learning

ABSTRACT

One of the common problems for L2 learners in ESL reading is that the text is either too difficult to understand or too simple to challenge, particularly the extracurricular reading material. *The Little Prince* and *Pride and Prejudice* are typically listed in the key reading materials for L2 learners. Contrary to the popularity, a large number of L2 learners find it is hard to complete due to the low text coverage with existing vocabulary. The corpus has become an important language research tool to store language materials according to the specific goals and the way through certain technical means. This paper is going to explore the text coverage of *The Little Prince* and *Pride and Prejudice* by range analysis, and comparing with English Curriculum Standards to figure out suitable reading targets of these two materials.

1. Introduction

The famous American linguist Wilkinson once said, without grammar, we can express very little; without words, we can express nothing. It can be seen that vocabulary is the basis of English learners' language ability. For many years, however, due to the influence of structural linguistics, English teaching has put too much emphasis on language structure and language form and neglected the role of vocabulary in L2 learning. When the vocabulary is not highly thought by language instructors, they would hardly consider the relationship between the vocabulary in teaching materials and L2 learners' existing L2 level.

There is no doubt that vocabulary is the foundation of understanding the reading materials. Klare (1963) pointed

out that the word is the most important unit in measuring readability and frequency^[1]. For second language teachers, choosing reading materials that are suitable for L2 learners' levels is extremely necessary. With the deepening of English education reform in China, in addition to English textbooks, extracurricular reading material like English novels has been progressively brought to light and attached importance.

The selected novels, *The Little Prince* and *Pride and Prejudice*, are generally featured in the best original editions recommended by English instructors in China. Influenced by the suggestion of expanding the amount of extracurricular readings, most L2 learners in lower levels would choose *The Little Prince* and in middle or advanced levels would choose *Pride and Prejudice*. However, most

*Corresponding Author:

Jiayu Zhuo,

Language Academy, University of Malaysia Technology, Malaysia;

Email: 2562275375@qq.com

learners face the problem of low text coverage with their existing English vocabulary and knowledge, which results in an easy abandonment of the original versions. By using BNC (British National Corpus), the list applied to see the word frequency, this article is to analyze vocabulary from the chapter 21 in *The Little Prince* and the chapter one in *Pride and Prejudice*, to see whether they are suitable for students who acquire the most frequent 800, 2400, 3500, and 4500 words. The ultimate purpose of this article is to provide methodologies of selecting appropriate reading materials as well as vocabulary course designing for instructors and provide methodologies to L2 learners when reading materials beyond their coverage.

2. Literature Review

Krashen's input hypothesis theory (1982) stated L2 learners can automatically output as long as getting enough input^[2]. Therefore, when low output abilities are brought to the forefront, input stage needs to be reflected as well, particularly materials selection. One of the criteria of measuring input material is vocabulary. As Lewis (1993) stated, the development of the four skills of a language depends on vocabulary learning, thus expanding L2 learners' vocabulary has been the central teaching goal of traditional ESL education^[3]. Since the 1980s, many scholars at home and abroad focus on textbook vocabulary and vocabulary has been a basic problem in the formulation of the syllabus and textbook compilation, as teaching materials are the main way to obtain knowledge. For Chinese English learners, the main way to learn vocabulary is teaching materials.

In the recent decades, globalization creates opportunities for more L2 learners to interact with English native speakers. Although communication is not a problem like years ago, L2 are still restricted to vocabulary and expressions learning from English textbooks in the output stage. In 1988, Sinclair and Renouf (1988) found that the key usage of some words presented in the textbook is very different from the real usage^[4]. Therefore, Chinese experts suggested authentic contexts should be encouraged, and original English reading material like novels, encourage ESL learners to do more extensive reading. For example, the English curriculum for senior middle schools (2017) states, in addition to teaching materials, the amount of extra-curricular reading needs to master over 230,000 words cumulatively to reach Level 7 and 300,000 to reach Level 8^[5]. Admittedly, even though extracurricular reading is helpful in improving English level, negative sides like merely pursuing difficult materials is severer than before, which may hurt L2 learners' confidence and cause unpleasant L2 learning feelings.

Regarding to it, many scholars explored the relationship between text coverage and reading comprehension. According to Nation (2006), with a text coverage of 95% (1 unknown word in 20), a few readers gained adequate comprehension, but they were still in the minority^[6]. Hu and Nation (2000) examined the relationship between text coverage and reading comprehension^[7]. They pointed out that the ideal unknown destiny was around 2%, meaning that 98% of the running words in the text should already be familiar to the learners. Carver (1994) indicates that even 98% coverage does not make comprehension easy^[8]. Kurnia (2003) found that few L2 learners gained adequate comprehension with 98% coverage^[9]. Laufer and Ravenhorst-Kalovski (2010) pointed out that it is useful to call 95% coverage minimal coverage and 98% adequate coverage^[10].

To test reading materials, Corpus provides abundant research samples for corpus linguistics by collecting a large number of authentic texts, and it provides new methods for English vocabulary teaching. The emergence of corpus linguistics makes it possible to assist daily English teaching and students' efficient learning. The paper is going to use this method to discuss the text coverage of extracurricular materials and discuss methodologies for ESL instructors regarding to extensive reading materials by using examples of *The Little Prince* and *Pride and Prejudice*.

3. Data Analysis

3.1 Analysis and Modification of *The Little Prince*

The English Curriculum in the whole basic education stage (including compulsory education and senior high school) sets nine levels according to the learners' English levels, forming a gradual and sustainable English learning procedure. According to English Curriculum Standards for Compulsory Education (2017-year edition), the total number of vocabulary in the compulsory stage is 2400^[11]. Level 2 is the basic requirement to meet at the end of grade 6, the required vocabulary of which is 800. Level 5 is the basic requirement that should be met at the end of grade 9, the required vocabulary of which is 1600.

As stated in the High School English Curriculum Standards, levels 6 to 9 are the objectives and requirements of ordinary high schools, and the total number of vocabulary is 4200. Level 7 is the basic requirement for high school graduation, the required vocabulary of which is 3500. Level 8 and level 9 are designed for high school students who are willing to further improve their English level. In the phase of post-secondary education, colleges and universities in China set the requirement of passing

CET 4 to graduate. In the syllabus for College English Test, the required vocabulary of CET 4 is 4500.

3.2 Analysis and Modification of *The Little Prince*

Table 1. Tokens, types and families at each word in chapter 21 in *The Little Prince*

| WORD LIST | TOKENS/% | TYPES/% | FAMILIES |
|------------------|-----------|-----------|----------|
| one | 793/86.95 | 218/78.70 | 174 |
| two | 38/ 4.17 | 27/ 9.75 | 24 |
| three | 24/ 2.63 | 9/ 3.25 | 9 |
| four | 28/ 3.07 | 7/ 2.53 | 6 |
| five | 3/ 0.33 | 1/ 0.36 | 1 |
| six | 14/ 1.54 | 5/ 1.81 | 3 |
| seven | 1/ 0.11 | 1/ 0.36 | 1 |
| eight | 0/ 0.00 | 0/ 0.00 | 0 |
| nine | 4/ 0.44 | 3/ 1.08 | 2 |
| ten | 1/ 0.11 | 1/ 0.36 | 1 |
| 11 | 1/ 0.11 | 1/ 0.36 | 1 |
| 12 | 1/ 0.11 | 1/ 0.36 | 1 |
| 13 | 0/ 0.00 | 0/ 0.00 | 0 |
| 14 | 0/ 0.00 | 0/ 0.00 | 0 |
| 15 | 0/ 0.00 | 0/ 0.00 | 0 |
| 16 | 3/ 0.33 | 2/ 0.72 | 2 |
| not in the lists | 1/ 0.11 | 1/ 0.36 | 0 |
| Total | 912 | 277 | 225 |

The data provided in Table 1 show the word coverage in chapter 21 from the book *The Little Prince*. According to the chart, this chapter has total 912 running words. The first row shows that the first 1,000 word-families from BNC account for 690 tokens, the coverage of which is 86.95%. The second row shows that the second 1,000 word-families from BNC account for 38 tokens, and these 38 words make up 4.17% of the total running words in the text. In addition, 24 of the running words in the text are in base list three and these 24 words make up 2.63% of the total running words in the text. Moreover, 28 of the running words in the text are in base list four and these 28 words make up 3.07% of the total running words in the text.

After detailed analyzing word list 16, it is obvious to see words “ah” and “oh” are all modal particles, and cannot influence the reading process. As for the word not found in the list, “grain fields”, it is a compound word. According to the chart, the minimum requirement to understand the text is mastering the most frequent 4000 words with the text coverage of 97.15%. However, students who master the most frequent 3000 words may find this chapter a little hard to read with 94.19% reading coverage, as they have not reached the minimal coverage: 95%. Preparing stage is needed for this group of students to read this chapter.

Connecting to general Chinese students’ situations, on

the one hand, to read this chapter as the extracurricular reading material, the most suitable group of students is university students and a small amount of high school students in Level 8-9, because they are capable of decoding information in the chapter with their existing ability.

On the other hand, high school students in Level 7 need to do some preparing actions before reading it such as searching background information. Tuinman and Brady (1974) found that enough pre-teaching resulted in little change in comprehension^[12]. Nation (2013) claims that if the vocabulary is an important factor in reading and readability measures, then pre-teaching the vocabulary that will occur in a reading text should increase the readability of the text^[13]. To begin with, with word list three, there is a word “prince”, occurring 14 times in this chapter. Similarly, the base list four shows that the word “fox” occurs 20 times in this chapter. For these two words, pre-teaching is an effective way to increase the readability of the text, since these two words are the most two important nouns in this chapter, showing two main characters of this chapter: a prince and a fox. Second, objects, pictures and so on are often considered as the most valid way of communicating the meaning of the word. Because these two words are nouns, teachers could use pictures or flash cards, helping students get the meaning of words. If students could understand these two words before they read the chapter, the text coverage would increase to 95.61%, reaching the minimal coverage. And this chapter becomes a little readable for those students. Third, with such coverage, if students still do not understand the contents, further efforts could be considered. In word list six, there are words “tame”, “tamed” and “tames”, occurring 8 times, 3 times and once respectively. Nation and Moir (2008) said when there are many unknown word families, replacing the unknown words with known or phrases which are likely to be understood by learners^[14]. Then, it is a good way for teachers to simplify the text by replacing the word family of “tame” by “control easily”, “controlled easily” and “controls easily” to reach the percentage of 96.93% since both word families of “control” and “easy” are in the base word list one and two in BNC. Beglar et al. (2012) found that the learners who made the most progress in the development of reading fluency over a year-long extensive reading program were not those who read the most, but those who read the most simplified material^[15].

Concluding analysis above, it is understandable that *The Little Prince* is an outstanding extracurricular reading material with its interesting contents, but chapter 21 is not suitable for ESL learners from Grade 1-8 in China, which means further data analysis of the remaining chapters are needed.

3.3 Analysis of *Pride and Prejudice*

Table 2. Tokens, types and families at each word in chapter one in *Pride and Prejudice*

| WORD LIST | TOKENS/% | TYPES/% | FAMILIES |
|------------------|-----------|-----------|----------|
| one | 690/88.47 | 236/77.38 | 189 |
| two | 38/ 4.87 | 30/ 9.84 | 29 |
| three | 8/ 1.03 | 7/ 2.30 | 7 |
| four | 6/ 0.77 | 5/ 1.64 | 5 |
| five | 4/ 0.51 | 4/ 1.31 | 4 |
| six | 2/ 0.26 | 2/ 0.66 | 2 |
| seven | 2/ 0.26 | 2/ 0.66 | 2 |
| eight | 0/ 0.00 | 0/ 0.00 | 0 |
| nine | 1/ 0.13 | 1/ 0.33 | 1 |
| ten | 0/ 0.00 | 0/ 0.00 | 0 |
| 11 | 1/ 0.13 | 1/ 0.33 | 1 |
| 12 | 0/ 0.00 | 0/ 0.00 | 0 |
| 13 | 0/ 0.00 | 0/ 0.00 | 0 |
| 14 | 0/ 0.00 | 0/ 0.00 | 0 |
| 15 | 20/ 2.56 | 10/ 3.28 | 10 |
| 16 | 1/ 0.13 | 1/ 0.33 | 1 |
| not in the lists | 7/ 0.90 | 6/ 1.97 | 0 |
| Total | 780 | 305 | 251 |

Table 2 provided above is the Chapter 1 from *Pride and Prejudice*. According to the chart, the running words of this chapter are 780 words. Based on the first two rows, the text coverage is 93.34%. Besides, the coverage in word list 15, the proper nouns list, is 2.56%. In word list 15, most vocabularies are persons' names like "Benny", "Bingley", "Jane", "Lizzy", "Lucas", "Lydia", "Morris" and "William", a holiday like "Michaelmas" or a place like "Chaise", which would not influence the procedure of reading. In addition to it, in the word list 16, words "oh" is a modal particle, which would not influence students' understanding as well. With these four lists, the text coverage of this chapter is 96.03%. Connecting to Chinese students' situations, students mastering the most frequent 2000 words (Level 6) or above are capable of reading this chapter as extracurricular reading material.

For students from Level 3-5, teachers need to make further efforts like giving glosses to students, making it appropriate for doing speed-training activities.

4. Reflections and Discussions

Based on the above data analysis, further reflections and discussion is going to be expanded regarding to the text coverage in ESL reading. Instructors need to give more attention to the text coverage of any materials during and after the class to assist L2 learners to improve language ability effectively and confidently. During the class time, when instructors do the part of presentation, direct and indirect vocabulary learning activities should

be involved. The most important principle in handling format and presentation is ensuring the vocabulary occurs across the four strands, namely meaning-focused input, language-focused learning, meaning-focused output and fluency development.

For example, in an extensive reading course, two strands should be considered^[13]. For the fluency development strand, reading materials should not contain any or very few unknown words because the goal is to read the book as quickly as possible in order to develop reading speed. For the meaning-focused input strand, the book should contain a few new words, but not too many, so that reading is not burdensome and there is the opportunity to learn new vocabulary incidentally while reading, making use of the guessing from context strategy. Carver (1994) argues that in meaning-focused language use, learners should know most of the language they need for a particular task, but a small percentage of words, about 1% to 5%, should be unfamiliar so that there is an opportunity for these items to be learned^[8].

If the teachers' aims are to do the meaning-focused training, teachers do not need to make further effort after modification for the Chapter 21 from *The Little Prince*. In terms of the Chapter 1 from *Pride and Prejudice*, they could directly provide the chapter to students and they do not need to anything to help students understand more unfamiliar words. However, if teachers focus on speed-training, it would be better to deal with more unfamiliar words for students^[13]. It is assumed that the highest density of glossing should be no more than 5%, and preferably around 3% of the running words. Therefore, taking the chapter from *Pride and Prejudice* as an example, this chapter could have around 24 to 39 glosses. With 24 more glosses, the text coverage would rise to more than 98%. Then this chapter could be used to do speed-training after adding glosses.

5. Conclusions

On a whole, there is no doubt that extensive reading is an effective way to improve language ability through 100% L2 input so that L2 learners' can output like native speakers. However, the precondition of selecting reading materials is the readability; otherwise, this method would easily lead to counter-productive result. The percentage of L2 learners' text coverage is one of the measuring tools of the readability, which is neglected by a large number of instructors.

One the one side, instructors need to create as many opportunities of immersing in L2 contexts as possible to make sure L2 learners receive enough input. One the other side, before requiring or recommending any original

materials, instructors should analyze the vocabulary in reading materials to make sure whether it is suitable for a certain level of students and design appropriate teaching plans to assist L2 learners in different levels so that they would benefit from the selected material to the maximum extent.

References

- [1] Klare, G. R. (1963). *The Measurement of Readability*. Ames, Iowa: Iowa State University Press.
- [2] Krashen, S.D. (1982) *The Input Hypothesis: Issues and Implications*. Longman, New York.
- [3] Lewis, B. (2010) The Lexical Approach [M]. *Language Teaching Publications*, London.
- [4] Sinclair, J. & A. Renouf. (1988). A Lexical Syllabus for Language Learning. In R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching* [C]. Harlow: Longman, 140-158.
- [5] The Ministry of Education of the People's Republic of China [M]. *English Curriculum Standards for Senior High School*. Beijing Normal University Publishing House, 2017.
- [6] Nation, P. (2006). How Large a Vocabulary is Needed for Reading and Listening? *The Canadian Modern Language Review*, 61, 1, 59-82.
- [7] Hu, M., & Nation, I. S. P. (2000). Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- [8] Carver, R. P. (1994). Percentage of Unknown Vocabulary Words in Text as a Function of The Relative Difficulty of the Text: Implications for Instruction. *Journal of Reading Behavior*, 26 (4), 413-437.
- [9] Kurnia, N. (2003). Returning of Multi-word Strings and Meaning Derivation from L2 Reading. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.
- [10] Laufer, B., & Ravenhorst- Kalovski. G. C. (2010). Lexical Threshold Revised: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, 22,1, 15-30.
- [11] The Ministry of Education of the People's Republic of China [M]. English Curriculum Standards for Compulsory Education. Beijing Normal University Publishing House 2017.
- [12] Tuinman, J. J., & Brady, M. E. (1974). How Does Vocabulary Account for Variance on Reading Comprehension Tests? A Preliminary Instructional Analysis. In Nacke, P. (ed.), *Interaction: Reading and Practice for College- Adult Reading*. Clemson, SC: National Reading Conference, 176-84.
- [13] Nation, P. (2013). *Learning Vocabulary in Another Language*. Cambridge University Press, United Kingdom.
- [14] Nation, I. S. P. & Moir, J. (2008). Vocabulary Learning and the Good Language Learner. In Griffiths, C. (ed), *Lessons from Good Language Learners*. Cambridge: Cambridge University Press, 159-73.
- [15] Beglar, D., Hunt, A., & Kite, Y. (2012). The Effort of Pleasure Reading on Japanese University EFL Learners' Reading Rates. *Language Learning*, 62, 3, 665-703.