The Analysis of Lexical Complexity of Two College English Textbooks

Zhen WANG

Foreign Language Department, Qingdao Agricultural University, Qingdao, China

Keywords: Lexical complexity, College English textbook, Lexical sophistication.

Abstract. The lexical complexity of two college English textbooks are analyzed in this article. Two sets of China’s Twelfth Five-Year National Planned College English Textbooks are selected as the research data, totally 28821 words. The Lexical Complexity Analyzer is applied to the whole lexical complexity and its trend of both textbooks. Generally speaking, the lexical density, sophistication and variation of Textbook B is higher than that of Textbook A. And there are different degrees of differences of texts of both textbooks in the above 3 measure categories of lexical complexity. Moreover, the trend lines of textbook B on the measures of lexical sophistication-II (LS2) and verb sophistication-II (VS2) increase more sharply than textbook A.

Introduction

As one of the elements that measure the characteristics of different discourses and stylistic features, Lexical complexity has always been a concern of researchers [1, 2, 3]. The previous researches mainly focus on the complexity of lexical use of English foreign language learners [4] and the measure methods of lexical complexity [5, 6]. Comparatively speaking, the research of lexical complexity of English textbooks has not been paid enough attention. Therefore, this study attempts to analyze the lexical complexity of college English textbooks.

The previous researchers put forward the different measurement index of lexical complexity. Generally speaking, the ways of measuring lexical complexity mainly focus on the following 4 aspects: the average length of a single word; the proportion of over-10-letter long words; the type-token ration of whole text and the different frequency magnitude of vocabulary use. Ai & Lu [2] at the Pennsylvania State University summarize 25 distinctive measurement index of lexical complexity (refer with: Table 1) and develop the automatic Lexical Complexity Analyzer (LCA) based on the previous researches. This tool makes the lexical complexity analysis automatic and simpler.

The Lexical Complexity Analyzer makes it possible to automatically analyze the lexical complexity of English textbooks. This will offer some reference and guidance for the objective evaluation of teaching materials, and provide some suggestions for the textbook selection.

Research Design

Research Questions

First, what is about the whole lexical complexity of two sets of textbooks? Second, does the lexical complexity of each textbook present the increasing trend? Third, are there differences between the lexical complexities of two textbooks?

Language Data

We selected two sets of textbooks as the language data in this study. Both textbooks are all China’s “the Twelfth Five-Year” National Planned Textbooks for higher education. Each set consists of 6 books, and each book consists of 8 units, and each unit consists of two texts. Usually, in China the college students are required to learn the first four books of one set of college English textbook and mainly study the first text of each unit, namely, text A, yet Text B is just for home reading. Eventually, each Text A of 64 units of 8 books of 2 sets of textbooks, totally 64 texts, are chosen as the language data for research. Moreover, 64 texts are organized in plain text format with the notes, charts, pictures and other auxiliary texts removed, and the words are 28821 in total.
Research Tool

The tool, LCA, allows language teachers and researchers to analyse the lexical complexity of written English language samples, using 25 different measures of lexical density, variation and sophistication proposed in the first and second language development literature \[3\]. The software runs on UNIX-like systems, and require the input texts to be part-of-speech (POS) tagged and lemmatized. This likely calls for familiarity of the command-line interface as well as some programming skills (e.g., part-of-speech tagging and lemmatization). We apply the web-based interface to LCA, available on this website (http://aihaiyang.com/software/lca/), to analyse the lexical complexity of textbooks. It eliminates the need for the command line interface and streamlines the above-mentioned natural language processing (NLP) processes, and generate the results in just a few clicks away. From According to LU \[3\], lexical variation involves 19 measures, and the main measurement index are the number of word types and the type-token ratios of words, particularly verbs, nouns, adjectives and adverbs. So we select “MSTTR, LV, CVV1, VV2, NV, ADJV, ADVV and MODV” 8 typical index as examples to analyse the lexical variation of texts. Finally we analyse 14 lexical complexity index of the texts in this study (refer with: Table 1).

Table 1. Measures of lexical density, sophistication and variation.

<table>
<thead>
<tr>
<th>Measure Types</th>
<th>Measures</th>
<th>Code</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical density</td>
<td>Lexical Density</td>
<td>LD</td>
<td>Nlex / N</td>
</tr>
<tr>
<td>Lexical Sophistication-I</td>
<td>LS1</td>
<td>Nslex / Nlex</td>
<td></td>
</tr>
<tr>
<td>Lexical Sophistication-II</td>
<td>LS2</td>
<td>T s / T</td>
<td></td>
</tr>
<tr>
<td>Verb Sophistication-I</td>
<td>VS1</td>
<td>Tsverb / Nverb</td>
<td></td>
</tr>
<tr>
<td>Verb Sophistication-II</td>
<td>VS2</td>
<td>Tverb /2Nverb</td>
<td></td>
</tr>
<tr>
<td>Corrected VS1</td>
<td>CVS1</td>
<td>T2sverb / Nverb</td>
<td></td>
</tr>
<tr>
<td>Lexical Variation</td>
<td>Mean Segmental TTR (50)</td>
<td>MSTTR</td>
<td>Mean TTR of all 50-word segments</td>
</tr>
<tr>
<td>Lexical Word Variation</td>
<td>LV</td>
<td>Tlex / Nlex</td>
<td></td>
</tr>
<tr>
<td>Corrected VV1</td>
<td>CVV1</td>
<td>Tverb / 2Nverb</td>
<td></td>
</tr>
<tr>
<td>Verb Variation-II</td>
<td>VV2</td>
<td>Tverb / Nlex</td>
<td></td>
</tr>
<tr>
<td>Noun Variation</td>
<td>NV</td>
<td>Tnoun / Nlex</td>
<td></td>
</tr>
<tr>
<td>Adjective Variation</td>
<td>ADJV</td>
<td>Tadj / Nlex</td>
<td></td>
</tr>
<tr>
<td>Adverb Variation</td>
<td>ADVV</td>
<td>Tadv / Nlex</td>
<td></td>
</tr>
<tr>
<td>Modifier Variation</td>
<td>MODV</td>
<td>(Tadj + Tadv) / Nlex</td>
<td></td>
</tr>
</tbody>
</table>

Results and Analysis

The Whole Lexical Complexity of Textbook A and B

Firstly, we apply the web-based LCA to analyse the lexical complexity of each text of each book and then import all values of lexical complexities of all texts of both textbooks (textbook A and textbook B) into the office software, EXCEL, to calculate the mean of the lexical complexities of 32 texts of textbook A and B respectively as the whole lexical complexity of both textbooks (refer with: Figure 1).

On the measure of lexical density, the value of LD of textbook B is higher than that of textbook A, 0.52 and 0.5 respectively. This indicates that the number of lexical words in textbook B is higher than that of textbook A, namely more lexical words in textbook B than in textbook A, such as verbs, nouns, adjectives and adverbs. On the measure of lexical sophistication-I, the value of LS1 of textbook B is higher than that of textbook A, 0.34 and 0.29 respectively, namely, more sophisticated lexical words among lexical words in textbook B than in textbook A. The sophisticated words, lexical words and verbs are those not on the list of the 2,000 most frequent words generated from the British National Corpus.

So the above data show that there are more infrequent words in textbook B than in textbook A. Likewise, on the measure of lexical sophistication-II, the value of LS2 of textbook B is higher than that of textbook A. This shows that there are more sophisticated word types among word types in textbook B than in textbook A. On the measures of VS1, VS2 and CVS1, the 3 values of textbook B are higher than those of textbook A. These 3 index illustrate that there are more sophisticated verbs
among all verbs in textbook B than in textbook A. Generally speaking, the values of lexical density and lexical sophistication of textbook B are higher than those of textbook A.

Figure 2 shows that all values of MSTTR, LV, CVV1, VV2, NV, ADJV and MODV of textbook B are slightly higher than those of textbook A, with the exception of ADVV, whose values are equal in both textbooks. On the measure of LV, there are more word types among words in textbook B than in textbook A, that is to say, the words are more rich and various in textbook B than in textbook A. On the measure of MSTTR, there are more types of lexical words among lexical words in textbook B than in textbook A, that is to say, the lexical words are more various in textbook B. On the measure of CVV1, there are more verb types among all verbs in textbook B. Namely, the verbs are more various and diversified in textbook B. Likewise, on the measures of VV2, NV, ADJV and MODV, there are more verb types, more noun types, more adjective types, more adjective + adverb types among lexical words in textbook B than in textbook A. What’s more, there are equal values of ADVV in both textbooks, namely, 0.07. This indicates that there are similar adverb type-lexical token ratio in both textbooks. Wholly speaking, there are much more lexical variation in textbook B than in textbook A.

The Trend of Lexical Complexity of Textbook A and B

We import all the values of lexical density, sophistication and variation of 32 texts of two textbooks respectively and obtain the following figures to illustrate the trend of lexical complexity of two textbooks.

From Figure 3, the value of lexical density of textbook A is between 0.44-0.55, and that of textbook B is between 0.47-0.59. This shows that there are different degrees of differences between 32 texts of each set of textbook. The degree of difference of lexical density of textbook A is 0.11 (0.55-0.44) and that of textbook B is 0.12 (0.59-0.47), which shows that there is almost the same degree of difference among texts of textbook A and B. However, Figure 3 illustrates that the trend lines of LD of both textbooks do not rise with the increasing numbers of texts of two textbooks. This can indicate that the lexical density does not increase as the number of texts of both textbooks increases.
From Figure 4, the value of lexical sophistication-II of textbook A is between 0.19-0.37, and that of textbook B is between 0.25-0.42. This shows that there are different degrees of LS2 between 32 texts of each set of textbook. The degree of difference of LS2 of 32 texts of textbook A is 0.18 (0.37-0.19) and that of textbook B is 0.17 (0.42-0.25), which shows that there is almost the same degree of difference among texts of textbook A and B. That is to say, there are the similar proportion of sophisticated word types among all word tokens in textbook A and B. However, Figure 4 illustrates that the trend lines of LS2 of both textbooks obviously rise with the increasing numbers of texts of two textbooks and the trend line of textbook B rises more sharply than textbook A. This can indicate that both LS2 of both textbooks increase as the number of texts increases, and textbook B is sharper than textbook A.

From Figure 5, the value of verb sophistication-II of textbook A is between 0.66-7.38, and that of textbook B is between 0.97-8.76. This shows that there are different degrees of VS2 between 32 texts of each textbook. The degree of difference of VS2 of 32 texts of textbook A is 6.72 (7.38-0.66) and that of textbook B is 8.59 (8.76-0.97), which shows that there is obviously different degree of difference of VS2 among texts of both textbooks. That is to say, there is much higher proportion of sophisticated verb types among all verb tokens in two textbooks. However, Figure 5 illustrates that the trend line of VS2 of textbook B rises more obviously with the increasing text numbers than textbook A, but the trend line of textbook A almost does not rise. This can indicate that the proportion of sophisticated verb types among all verb tokens in textbook B increases sharply, not in textbook A.
From Figure 6, the value of MSTTR of textbook A is between 0.7-0.83, and that of textbook B is between 0.77-0.83. This shows that there are different degrees of MSTTR between 32 texts of each set of textbook. The degree of difference of MSTTR of texts of textbook A is 0.13 (0.83-0.7) and that of textbook B is 0.06 (0.83-0.77), which shows that there is much more different degree of difference of MSTTR among texts of textbook A than that of textbook B. That is to say, the words are more various and diversified in textbook A than textbook B. What’s more, Figure 6 illustrates that the trend line of MSTTR of textbook B almost does not rise with the increasing numbers of texts, yet the trend line of textbook A decrease with the increasing numbers of texts. That’s to say, the words do not become more various in textbook B, but more undiversified in textbook A.

From Figure 7, the value of LV of textbook A is between 0.45-0.81, and that of textbook B is between 0.53-0.85. This shows that there are different degrees of LV between 32 texts of each set of textbook. The degree of difference of LV of texts of textbook A is 0.36 (0.81-0.45) and that of textbook B is 0.32 (0.85-0.53), which shows that there is almost the same degree of difference of LV among texts of textbook A and B. That is to say, the proportion of lexical word types and lexical words is almost the same high in textbook A and B. What’s more, Figure 7 illustrates that both trend lines of LV of textbook A and B rise with the increasing numbers of texts. That’s to say, the lexical words types become more various in both textbooks.

Conclusion

In general, the lexical density, sophistication and variation of texts of textbook B are higher than that of textbook A. Specifically, on the measure of lexical density (LD), there exists the different degrees of difference among the texts in both textbooks, but both trend lines of two textbooks do not increase. On the lexical sophistication-II (LS2), there is almost the same degree of difference among texts of both textbooks, and the trend lines of LS2 of both textbooks obviously rise, but the trend line of textbook B rises more sharply than textbook A. On the verb sophistication-II (VS2), there are different degrees between texts of both textbooks, and the proportion of sophisticated verb types in textbook B increases sharply, but not in textbook A. On the mean segmental TTR (50) (MSTTR), there are different degrees of difference among texts in both textbooks, and the MSTTR almost does not increase, but it decreases in textbook A. On the lexical word variation (LV), there are different degrees of difference in both textbooks, and the trend lines of LV of both textbooks increase. It is hoped that this study will offer some implications for the future textbook difficulty evaluation. And it is very useful to make the similar analysis for the selection of suitable teaching materials to students.

Acknowledgement

This research was financially supported by Shandong Educational Science “13th Five-Year” 2016-2017 Planning project (project number: BCGW2017001).
References


