Word number, level, and frequency in current English course books: Corpus study

Ari Nurweni 1, Endang Komariah 2
1 English Education Department, Teacher Training and Education Faculty, Universitas Lampung, Lampung, Indonesia
ari.nurweni@fkip.unila.ac.id
2 English Education Department, Teacher Training and Education Faculty, Universitas Lampung, Lampung, Indonesia
endang.komariah@fkip.unila.ac.id

DOI: http://dx.doi.org/10.26418/jeltim.v1i1.64468

Abstract

Introducing the right English words in an adequate number and levels to beginner EFL learners is of paramount importance to enable them to communicate and develop their English further. This corpus study is meant to investigate the number and levels of English words, and their frequency of occurrences in a set of three official English course books for Indonesian junior high school students of the seventh to ninth grades, published by The Indonesian Ministry of Culture and Education. The study was conducted by converting each of the three English books into MS Word documents, eliminating the non-English and very common loanwords, saving each course book in plain text, and processing the plain text using a word frequency analysis program. The results show the English course books contain (1) a sufficient number of English words for comprehension of simplified pictured texts but not for authentic texts, (2) the high-frequency words that belong to the first, second thousand, and academic words, each of which is in an inadequate number, and (3) the frequency of word occurrences in the three books that is unlikely to lead to incidental word learning if no further efforts are made by the students and teachers.

Keywords: course book; English word; frequency; word family


To link to this article: http://dx.doi.org/10.26418/jeltim.v1i1.64468
The number of English words is numerous. For example, Collins English Dictionary contains 722,000 entries (https://blog.collinsdictionary.com/language-lovers/collins-english-dictionary-12th-edition-faqs/). Webster’s Third New International Dictionary, Unabridged, contains 470,000 entries (https://www.merriam-webster.com/help/faq-how-many-english-words). The Oxford English Dictionary contains around 290,500 main entries (https://public.oed.com/history/oed-editions/introduction-to-the-second-edition/). It is almost impossible to present all of the words in a limited learning time in a class. So, which English words need to be introduced to EFL young learners as scaffolding to study English further? English words they get from the class should help them understand English texts that they read and listen to. Fata et al. (2022) confirm that the number of words in a sentence and the number of syllables in a word will determine the level of difficulty of an English text. Moreover, if there are many unknown words in a sentence, the sentence will be difficult to understand, and this will not enable the readers to guess the meaning of the unknown words in the sentence correctly.

Nation (2001, pp. 18-20) classifies English words into high-frequency words, mid-frequency words, low-frequency words, academic words, and technical words. The high-frequency words are those which appear very frequently in any English text. These words are grouped into the first and second thousand words. The first thousand words consist of 986 word-families with 4119 word-types and the second thousand most frequent words consist of 988 word-families with 3708 word-types. The first and the second thousand words consisting of 1974 word families (coming from General Service List or GSL) cover 84.24% of The Cambridge English Corpus (CEC) and 67.64% of the New General Service List (NGSL). NGSL consisting of 2368 word families with 2809 word types covers 90.34% of CEC (Browne, 2013). In fact, 66.64% of word types of the NGSL belong to the GSL, 19.65% to academic words created by Coxhead (2000), and 12.71% are not on the lists. Thus, when the words of the GSL and the academic words are combined, their coverage in the CEC will be very close to that of the NGSL in CEC.

Academic words appear very frequently in academic texts across disciplines. Academic words, consisting of 570 English word families with 3,107-word types, cover 10% of the words in academic texts (Coxhead, 2000) and 11.95% in the abstracts of scientific journal articles (Jemadi et al., 2019). In addition, Gholaminejada and Sarab (2021), based on others’ study findings, state that academic word mastery also plays a significant role in students’ reading and writing proficiency. Syarofi and Shobaha (2023), reviewing 18 studies on the effects of academic word list learning on students’ writing achievement, found that 15 show academic word list learning improves students’ writing.
Technical words frequently occur in a specialized text. The number of technical words usually ranges from 1,000 to 2,000 for each specialized field. The technical words will be learned best when learners study the subject matter. Whereas, low-frequency English words appear very rarely in any English text.

In addition to the number and level of English words in course books, the word frequency of occurrence also needs to be considered as it contributes to word learning (Aldukhayel, 2022; Altalhab, 2018; Joseph & Nation, 2018; Brysbaert et al., 2018). Aldukhayel (2022) conducted a comparative study of 63 EFL learners to see the effect of having words with three or more appearances in a documentary video with subtitles to a group of students and having words with just one appearance in a video they watched. The findings give evidence that word frequency of occurrence affects incidental word learning. Joseph and Nation (2018) and Brysbaert et al. (2018) come to a similar conclusion even though word frequency is not the only factor that leads to the words learned. Brysbaert et al. further state that word frequency is highly associated with a number of other word features, such as word length, age at which the word was acquired, and resemblance to other words; basically, the word frequency effect is likely to be combined with any of these factors.

Several studies, namely Hardianti and Sudarsono (2021), Prihatiyningsih, Petrus, and Silvhiany (2021), Rahmawanti et al. (2019), Fuyudloturromaniyyah (2015), Aziz and Aziez (2018), Ikhsanudin & Suparjan (2018), and Rahmany and Khan (2022), have investigated English course books for Indonesian learners from different angles. Hardianti and Sudarsono (2021) studied an English course book for grade seven ‘Bahasa Inggris When English Rings a Bell’. Their study is focused on the variety of English. The finding reveals that the course book contains American, British, and American and British-like varieties in words, spelling, and grammar.

Prihatiyningsih, Petrus, and Silvhiany (2021) studied the cultural representation of the English course book for grade 7. Another study by Rahmawanti et al. (2019) evaluates the quality of an English course book for grade 8 junior high school students, ‘Bahasa Inggris When English Rings a Bell’. The study revealed that the course book fits good criteria regarding aims and approaches, language skills, and topics, but it needs to fit the design, the organization, and the language content. From the teachers’ view, the course book is considered credible to be used by the learners and almost all materials in the course book are consistent with basic competencies 3 and 4 of Curriculum 2013.

Fuyudloturromaniyyah (2015) investigated three chapters (1, 6, and 11) of the selected course book for grade 7, Bahasa Inggris When English Rings a Bell. The texts were analyzed by using a task analysis sheet. The findings showed that in terms of physical appearance and types of activities, the course book was attractive since it is printed in full colour and contains pictures, illustrations, different types of exercises and tasks, songs, and games, but the activities do not
reflect that the stages of the scientific approach as prescribed by Curriculum 2013 are not done in balanced proportions.

Aziz and Aziez (2018) examined the English vocabulary that appears in English course books for junior high school students, entitled English in Focus, and for high school students entitled Developing English Competencies and those that appear in the national exam questions for junior high and senior high school students. In fact, school books for Indonesian students have been assessed in terms of content, methodology, language, and graphic design with a rigid standard (Ikhsanudin & Suparjan, 2018). Rahmany and Khan (2022) evaluated English textbooks for Afghanistan students. Their study focused on the organization of the books and the adequacy of sub-topics in each unit but did not analyze the number, levels, and frequency of the English words in the textbooks.

Based on the literature review above, none of the previous studies on English course books investigates the English words in a full set of official English course books for grades seven, eight, and nine of Indonesian students. Furthermore, Indonesian EFL learners’ English vocabulary mastery still needs to improve (Nurweni, 2017 and Novianti, 2016). Accordingly, this study aims to answer: (1) Is the total number of English words in the three English course books sufficient for Indonesian students of the seventh to ninth grades to read English texts?; (2) Are the numbers of the first thousand, second thousand, and academic words in the course books adequate?; and (3) Does the frequency of occurrence of the English words lead to incidental word learning in the course books?

The answers to these research questions are expected to be able to shed light on how to help learners make use of the books most effectively so that the learners can increase their English word mastery optimally. Knowing the vocabulary description of the books is very important as a basis for developing teaching and learning materials and methods in order to help the students achieve optimum results in learning English. It will also provide authors of English course books with a basis create English course books that introduce English words in an adequate number, at the right level, and sufficient frequency of occurrences.

METHOD

This research is a corpus study, analyzing the number, level, and frequency of English words in a set of English course books for Indonesian junior high school (JHS) students. The analysis covers word list, token, type, and word family, as well as the frequency of word occurrences in the course books. Therefore, the English words in a set of English course books for JHS that consist of English course books for grades 7, 8, and 9 are investigated. The titles of the books are When English Rings a Bell for Grade 7 written by Siti Wachidah, Asep Gunawan, Diyantari, Yuli Rulani Khatimah (revised in 2017), Bahasa Inggris, When English Rings a Bell for Grade 8 by Siti Wachidah, Asep Gunawan,
Diyantari, Yuli Rulani Khatimah (written in 2017), and Think Globally Act Locally for grade 9 by Siti Wachidah, Diyantari, and Yuli Rulani Khatimah (written in 2015). The course books were studied because they were official course books and published by the Indonesia’s Educational and Cultural Ministry. In addition, the books are used by Indonesian junior high school students and teachers up to now, meaning that the books will have an impact on many students nation-wide.

The data collection stages emulated those written in the manual of the RANGE program, obtained from http://www.vuw.ac.nz/lals/staff/Paul_Nation. The RANGE program was designed by Paul Nation and Averil Coxhead of Victoria University, Wellington, New Zealand, and was programmed by Alex Heatley. The first version of the Range program is called Fvords and VocabProfile. The RANGE program was selected because it can perform analyses for this study. The stages of collecting the data included the followings. The first was converting the texts of the course books into MS Word file. The second was removing non-English words, proper nouns, words belonging to metric units, and images from the texts. Proper nouns and English words of metric units were excluded because they are borrowing words that students were very familiar with. The third was removing unnecessary spaces between words in the texts that appeared due to the file format conversion. The fourth was saving the edited MS Word texts as plain texts for each of the course books.

The data analysis stages to address research objectives 1 and 2 were: (1) analyzing the English words of the course books which have been saved in three different plain text files using the RANGE program; (2) checking and editing the words that fall into the category of not on the list to see if there are non-English words, and the words not on the list are put into words families manually since the program does not have a feature that can do that; and (3) putting the summary statistics in tables. We used the RANGE program to count words as tokens, word types, and word families of the first and second thousand most frequent words, and academic words. Tokens are any single words in spoken or written running texts. For example, the sentence: “The speaker is speaking to another speaker in the room” consists of 10 tokens. Word types are words that have different forms so that sentence consists of eight word types. Word families are words that have the same base words and are counted as one word family. For example, “speaker” and “speaking” are counted the same word family of “speak”. Therefore, the example sentence above consists of seven word families.

To address research objective 3, we conducted three-stage analysis. This involved: (1) reorganizing the word lists resulting from the analysis using the RANGE program based on the frequency by range, type, and word family; (2) classifying the words based on whether they occur only in one book, two books or in the three books; and (3) grouping the words in the lists based on their frequency of occurrence in the course books.
FINDINGS

The findings of this study are grouped into: (1) the number of English words in three related English course books; (2) the number of English words by word list, token, type, and word family in each of the course books; and (3) the frequency of word occurrence in the three course books and in each of the course books.

1 The number of English words in three English course books

The words in the three English course books were analyzed simultaneously to get the total number of English words in each of the course books for the Indonesian students of the seventh to ninth grades and in three books as a whole. The results of the analysis are displayed in Table 1 below.

Table 1. The number of English words in the three English course books

<table>
<thead>
<tr>
<th>English Course Books</th>
<th>Word Tokens</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 7</td>
<td>20,684</td>
<td>1,312</td>
<td>992</td>
</tr>
<tr>
<td>Grade 8</td>
<td>20,198</td>
<td>1,432</td>
<td>1,008</td>
</tr>
<tr>
<td>Grade 9</td>
<td>29,566</td>
<td>2,195</td>
<td>1,506</td>
</tr>
<tr>
<td>Grades 7, 8, and 9 as a whole</td>
<td>70,448</td>
<td>3,050</td>
<td>1,988</td>
</tr>
</tbody>
</table>

Table 1 shows that the numbers of tokens, word types, and word families for each of the course books are different from the book for grades seven to nine. The book for grade seven contains the lowest word types, 1,312, and word families, 992. On the other hand, the book for grade nine contains significantly the highest number of word types, 2,195, and word families, 1,506. When words in the three books are counted as a whole, the numbers of word types and word families are 3050 and 1988 respectively. These numbers are smaller, not the same as the total numbers of words in each of the books for grades 7, 8, and 9. This implies that around 1889 word types and 1,518 word families appear in more than one book.

The total numbers of words in the three books counted as a whole as word type and word family indicate that when the students use the three books until they finish their study at junior high school, they will likely know 3,050 word types and 1,988 word families. These figures are sufficient for the students to learn and enable them to read simplified pictured English texts (Nation, 2001). Learners need 8,000 to 9,000 English word families to understand authentic English texts and 6,000 to 7,000 ones to comprehend spoken English texts (Nation, 2006). However, van Zeeland and Schmitt (2012) show that ones need a lower number, that is, 2,000 to 3,000 English words for listening comprehension.
2 The Levels of English words in three English course books

The levels of English words in the three course books are classified by using the word lists proposed by Nation (2001). Table 2 below displays that 1461 (47.90%) of the 3050 word types belong to the first thousand most frequent words, 685 (22.46%) belong to the second thousand most frequent words, 172 (5.64%) belong to the academic word list, and 732 (24.00)% belong to the words that are not in the other three word-lists.

Table 2. The Number of English Words in the Three English Books by Word List

<table>
<thead>
<tr>
<th>Word List</th>
<th>Word Tokens</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>First thousand words</td>
<td>58,745 (83.31%)</td>
<td>1461 (47.90%)</td>
<td>761 (38.24%)</td>
</tr>
<tr>
<td>Second thousand words</td>
<td>5,571 (7.9%)</td>
<td>685 (22.46%)</td>
<td>472 (23.72%)</td>
</tr>
<tr>
<td>Academic words</td>
<td>1,066 (1.51%)</td>
<td>172 (5.64%)</td>
<td>129 (6.48%)</td>
</tr>
<tr>
<td>Not on the lists</td>
<td>5,066 (7.19%)</td>
<td>732 (24.00%)</td>
<td>626 (31.56%)</td>
</tr>
<tr>
<td>Total</td>
<td>70,448 (100%)</td>
<td>3050 (100%)</td>
<td>1,988 (100%)</td>
</tr>
</tbody>
</table>

Table 2 shows that by token, the numbers of the first thousand most frequent words, the second thousand most frequent words, the academic words, and the English words not found in the lists are 58,745, 5,571, 1,066, and 5,066 respectively, accounting for 83.31%, 7.9%, 1.51%, and 7.19% respectively. By word family, the numbers of the first thousand most frequent words, the second thousand most frequent words, the academic words, and the English words not found in the lists are 761, 472, 129, and 626 respectively, accounting for 38.24%, 23.72%, 6.48%, and 31.56% respectively. It can be inferred from Table 2 that the coverages of the first thousand, second thousand, and academic words in the three course books are 92.81% by token, 76 % by type, and 68.44% by word family.

Table 3 displays the number of English words in the English course book for grade 7 by word list, token, type, and family.

Table 3. The number of English words in the course book for grade 7 by word list

<table>
<thead>
<tr>
<th>Word Lists</th>
<th>Word Tokens</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>First thousand words</td>
<td>17,280 (83.54%)</td>
<td>748 (57.01%)</td>
<td>515 (51.92%)</td>
</tr>
<tr>
<td>Second thousand words</td>
<td>1,643 (7.94%)</td>
<td>273 (20.81%)</td>
<td>216 (21.77%)</td>
</tr>
<tr>
<td>Academic words</td>
<td>243 (1.17%)</td>
<td>51 (3.89%)</td>
<td>42 (4.23%)</td>
</tr>
<tr>
<td>Not on the lists</td>
<td>1,519 (7.35%)</td>
<td>240 (18.29%)</td>
<td>219 (22.08%)</td>
</tr>
<tr>
<td>Total</td>
<td>20,685 (100%)</td>
<td>1312 (100%)</td>
<td>992 (100%)</td>
</tr>
</tbody>
</table>
Table 3 shows that the English course book for grade 7 contains 20685 tokens, consisting of 1,312 word types, and 992 word families. Of the 1,312 word types, 748 (57.01%) belong to the first thousand most frequent words, 216 (21.77%) belong to the second thousand most frequent words, and 42 (4.04%) belong to the academic word list, while 219 (22.08%) belong to the words that are not in the word lists. It can be inferred from Table 3 that the coverages of the first thousand, second thousand, and academic words are 92.65% by token, and 81.71% by type, 77.92% by word family in the English course book for grade 7.

Table 4 shows that the English course book for grade 8 contains 20199 tokens, consisting of 1,432 word types. Of the 1,432 word types, 882 (61.59%) belong to the first thousand most frequent words, 302 (21.09%) to the second thousand most frequent words, and 52 (3.63%) to the academic word list, and 196 (13.69%) to the words that are not in the three word lists. It can be inferred from Table 4 that the coverages of the first thousand, second thousand, and academic words are 94.76% by token, 86.31% by word type, and 82.44% by word family in the English course book for grade 8.

Table 5 shows that the English course book for grade 9 contains 29,569 tokens, consisting of 2,186 word types. Of the 2,195 word types, 52.77% of them belong to the first thousand most frequent words, 20.77% to the second thousand most frequent words, and 6.01% to the academic word list, while 20.64% belong to the
words that are not in the other three word lists. It can be inferred from the table that the coverage of the first thousand, second thousand, and academic words is 91.51, by token, 79.36% by type, and 74.30% by word family in the English book for grade 9.

Most of the English words in the three course books, 76 % by word type, and 68.44% by word family belong to those that are needed to learn and master by the students. Those words are right to be introduced to junior high school students. They frequently appear in any English texts and their coverage in English texts is very high (Nation, 2006). However, the numbers of the first thousand, second thousand, and academic words are not adequate in the course books. It is called adequate if the number reaches 80% of the words in each of the word list.

Concerning with the number of the English words that are off the lists that appear in the course books, it can be said that its percentage exceeds that usually appears in authentic English texts as stated by Nation (2001). These words are needed in the course books but they are rarely needed in their daily life and so they do not need to be introduced in the early stage of learning English as a foreign language as they do not belong to the high frequency words. The learners will rarely meet these words in authentic English spoken and written texts they will have in their life.

3 Frequency of word occurrences in English course books for JHS

The frequency of occurrences of English words in the three English books for the JHS students is viewed from 3 sides. The first one is the range of the word occurrence, that is, whether the words appear in one, two, or three English course books (Table 6).

Table 6. Number of word families occurring in the three English course books

<table>
<thead>
<tr>
<th>Word families appearing in</th>
<th>First Thousand</th>
<th>Second Thousand</th>
<th>Academic Words</th>
<th>Words not on the Lists</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>one book</td>
<td>188 (24.70%)</td>
<td>247 (52.33%)</td>
<td>89 (68.99%)</td>
<td>507 (81.05%)</td>
<td>1,031 (51.91%)</td>
</tr>
<tr>
<td>two books</td>
<td>162 (21.29%)</td>
<td>133 (28.18%)</td>
<td>22 (17.05)</td>
<td>84 (13.38%)</td>
<td>401 (20.15%)</td>
</tr>
<tr>
<td>three books</td>
<td>411 (54.01%)</td>
<td>92 (19.49%)</td>
<td>18 (13.95)</td>
<td>35 (5.57%)</td>
<td>556 (27.94%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>761 (100%)</td>
<td>472 (100%)</td>
<td>129 (100)</td>
<td>626 (100%)</td>
<td>1,988 (100%)</td>
</tr>
</tbody>
</table>

Table 6 displays that 1,031 (51.91%) of the word families occur only in one of the three English course books, 401 (20.15%) in two of the books, and 556 (27.94%) in the three books. The table also shows that 411 (54.01%) of the word families in the first thousand occur in the course books. On the other hand, most of the second thousand, academic words, and words not on the lists appear only in one of the course books.
The second one is the frequency of occurrence of the English words in the set of three course books is counted simultaneously, and the third is the frequency of occurrence of English words in each of the three books is counted separately as presented in Table 7.

Table 7. Word Occurrence Frequencies in the Course Books

<table>
<thead>
<tr>
<th>Word Lists</th>
<th>Times of Occurrence</th>
<th>In Grade 7 Book</th>
<th>In Grade 8 Book</th>
<th>In Grade 9 Book</th>
<th>In Grade 7, 8, &amp; 9 Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>First thousand words by word family</td>
<td>5 ≤ 291 (56.50%)</td>
<td>300 (53.96%)</td>
<td>432 (64.09%)</td>
<td>476 (70.21%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-4 133 (25.80%)</td>
<td>169 (30.39%)</td>
<td>150 (22.26%)</td>
<td>116 (17.11%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1     91 (17.70%)</td>
<td>87 (15.65%)</td>
<td>92 (13.65%)</td>
<td>86 (12.68%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>515 (100%)</td>
<td>556 (100%)</td>
<td>674 (100%)</td>
<td>678 (100%)</td>
<td></td>
</tr>
<tr>
<td>Second thousand words by word family</td>
<td>5 ≤ 59 (27.31%)</td>
<td>63 (26.92%)</td>
<td>103 (30.21%)</td>
<td>177 (37.50%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-4 74 (34.26%)</td>
<td>86 (36.75%)</td>
<td>130 (38.12%)</td>
<td>177 (37.50%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1     83 (38.43%)</td>
<td>85 (36.32%)</td>
<td>108 (31.67%)</td>
<td>118 (25.00%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>216 (100%)</td>
<td>234 (100%)</td>
<td>341 (100%)</td>
<td>472 (100%)</td>
<td></td>
</tr>
<tr>
<td>Academic words by word family</td>
<td>5 ≤ 11 (26.19%)</td>
<td>16 (39.02%)</td>
<td>25 (24.04%)</td>
<td>44 (34.38%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-4 16 (38.10%)</td>
<td>10 (24.39%)</td>
<td>43 (41.35%)</td>
<td>42 (32.81%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1     15 (35.71%)</td>
<td>15 (36.59%)</td>
<td>36 (34.62%)</td>
<td>42 (32.81%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42 (100%)</td>
<td>41 (100%)</td>
<td>104 (100%)</td>
<td>128 (100%)</td>
<td></td>
</tr>
<tr>
<td>Words not in the lists by word family</td>
<td>5 ≤ 19 (8.64%)</td>
<td>19 (10.73%)</td>
<td>68 (17.52%)</td>
<td>116 (18.53%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-4 108 (49.54%)</td>
<td>74 (41.81%)</td>
<td>165 (42.53%)</td>
<td>259 (41.37%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1     92 (41.82%)</td>
<td>84 (47.46%)</td>
<td>154 (39.95%)</td>
<td>251 (40.10%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>219 (100%)</td>
<td>177 (100%)</td>
<td>387 (100%)</td>
<td>626 (100%)</td>
<td></td>
</tr>
<tr>
<td>Grand Total</td>
<td>992 (100%)</td>
<td>1,008 (100%)</td>
<td>1,506 (100%)</td>
<td>1,988 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 7 shows that when the books were analyzed one by one, the percentages of the first thousand most frequent words that appear 5 times or more in English course books for grades 7, 8, and 9, range from 50% to 64%, of the second thousand most frequent words, and academic words range from 24.04% to 39.02%, and of the words not found in the lists range from 8.64 % to 17.52%. When the three books were analyzed simultaneously, the results are similar, that is, the number of the first thousand most frequent words appearing five times and more is the highest, 476, accounting for 70.21%. Table 7 above also shows that the percentages of the words in all the word categories appearing less
than five times in each of the three books range from 13.65% to 49.54%, and range from 12.68% to 41.37% when the three books were analyzed simultaneously.

The examples of the ten most frequent first thousand words found in the three English course books are the, we, be, will, to, and, I, in, you, and a. They appear at least 261 times in the course books. Most of the words belong to functional words, such as articles, pronouns, prepositions, and coordinators. The functional words whose frequency of occurrence is one (very low) are among others, among, except, and within. The content words belonging to the first thousand word-families with a frequency of occurrence one are such as able, account, belong, church, common, gentle, iron, quite, rise, wound, and youth.

The examples of the ten most frequent second thousand words appearing in the three course books are correct, sentence, loud, conversation, spell, dictionary, inform, repeat, copy, and collect. They appear at least 98 times in the course books. The words are associated with things needed in a class and activities performed in class. The words are content words that belong to verbs, nouns, and adjectives. The examples of the second thousand word-families whose frequency of occurrence is one and only appear in one of the books are throat, thumb, tip, toe, tonight, upset, vowel, warn, weigh, and wheel.

The examples of the ten most frequent academic word-families in the three English course books when the words in the three books were processed simultaneously are role, chapter, final, text, communicate, journal, analyze, identify, locate, and label. They appear in the course books at least 21 times in total. The following are examples of the ten least frequent academic word families in the three English course books when the words in the three books were processed simultaneously: proceed, prohibit, recover, resource, reverse, revise, summary, temporary, tense, and vary. These academic words appear once in only one of the three books.

The examples of the ten most frequent words not found in any other word lists in the three English course books when the words in the three books were processed simultaneously are punctuation, handwrite, orally, classroom, notebook, orphan, buffalo, lyrics, homework, and magazine. Those words appear at least 21 times in the three books in total. The words orphan and buffalo appear 27 and 24 times respectively but in only one of the books. On the other hand, the examples of the ten least frequent words not found in any other word lists in the three English course books when the words in the three books were processed simultaneously are trespassing, troops, tyre, uniforms, ward, wardrobe, wavy, worksheet, yoghurt, and zookeeper. These words appear once in only one of the three books.

The data show that most of the English words appear only in one of the three course books and appear less than six times. This means that the word frequency of occurrences in the course books is unlikely to lead to incidental word learning. Nation (2015) explains that there are two things that lead to word learning and acquisition, that is, the number of word appearances in spoken and
written texts and the quality of students’ attention to the words. Nation further explains that to know a word deeply the word must appear many times in different contexts. Schmitt et al. (2017) show that meeting a word for eight to ten times is needed to know the word well.

DISCUSSION

The first finding of this study is that the numbers of tokens, word types, and word families in the three course books for grades seven to nine are 70,448, 3,050, and 1,988 respectively. The number of tokens is much bigger than that of the word types and the word family when the words in the three books were analyzed at once and when each was analyzed separately. This is due to the fact that there are many English words in the course books that appear more than one time. As defined in the previous part, tokens are any words, whether they have the same forms or different forms, they will be counted as words, called tokens. The number of words in terms of word type is smaller than the number of tokens but greater than the word families. As it is explained in the previous part, all words, even though they have different word types, if they belong to the same base word, are counted as a word family. Therefore, the number of word families is the smallest. It is also observed that the number of tokens, word types, and word families gradually increases from lower grades to the highest grade. This indicates that the English words in the course books are properly developed in line with the learning capacity of the learners. In addition to that, the course books for grades seven and nine, but not for grade eight, in this study contain greater numbers of tokens and word types than those in other JHS books (English in Focus) studied by Aziez and Azies (2018).

When the words in the three books were counted as a whole, the numbers of word types and word families were smaller than those counted separately in each of the course books, and then each of the word type numbers and the word family numbers in each of the books were totalled. This is because around 38% of the word types and 21% of the word families are found in more than one of the course books. The numbers of word types and families are adequate for the EFL students to read pictured English texts and communicate in spoken daily English (Nation, 2001). Conversely, those numbers are not sufficient to enable them to comprehend authentic English text. EFL learners need 8,000 to 9,000 English word families. Whereas, EFL learners need 6,000 to 7,000 English words for listening comprehension (Nation, 2006). Previous studies show that there are significant positive correlations between one’s English vocabulary mastery and
reading comprehension skills (Gu, 2017; Moon, Choi, & Kang, 2019; Kang, 2020, and Park & Chon, 2018).

The second finding shows that the three English course books contain four different levels of English words as classified by Nation (2001), namely the first thousand most frequent words, the second thousand most frequent words, academic words (the words in these three lists are called high-frequency words), and low-frequency words (not in the other three lists). By token, the first thousand most frequent words account for the highest percentage (83.31%), followed by the second thousand most frequent words (7.9%), the English words not found in the lists (7.9%), and the academic words (1.51%). The percentage is similar to that estimated by Nation (2001). Furthermore, the coverage of the first and second thousand most frequent words is also in line with those found in the CBC by Browne (2013) and in an English course book for grade 12 studied by Selamat (2018). The coverage is normal as the first and second thousand most frequent words frequently appear in any English texts; moreover, the course books are for EFL beginners.

The numbers of the first thousand and the second thousand-word families in each of the course books also increase significantly from one grade to another grade. However, by word family, the numbers of the first and second thousand most frequent words that appear in the course books are not at the maximum numbers of the first thousand and second thousand most frequent words, that is, 986 and 988, respectively. This is partly because the number and size of course books for learners are limited, and the topics and the exercises are also limited so they can cover some of the high-frequency words to appear in the course books. In addition, in some cases, the authors select low-frequency words instead of the high-frequency ones that are synonymous with the low-frequency words selected.

Concerning academic words, Coxhead (2000) lists 570 academic words that frequently appear in English academic texts. However, the English course books analyzed contain a small number of academic words, 1,066 by token, 172 by word type, and 129 by word family. It is less than one-fourth of that number of academic words, even though the number of academic words from grade seven to grade nine significantly increases. Low coverage of AWL is also found by Selamat (2018), even in an English course book for grade 12. The low coverage of AWL in English course books for grades seven to nine is likely because the course books are intended for EFL young learners. The goals of learning English for them are mainly to be able to communicate in the daily English language and to understand simple functional English texts such as invitation cards, food labels,
announcements, and simple monologue texts such as descriptive, procedure, and recount texts. Therefore, the number of academic words in the course books is small.

The English course books for students of grades seven to nine contain a significant number of non-high-frequency English words (not in the lists). By token, the number is ideal and in line with the number of words not in the word lists in authentic English texts, but by word type and family, the numbers are not ideal, exceeding the figures in authentic English texts (Nation, 2001). There are several reasons why they appear in the course books. Firstly, they are used to explain the language the students learn in the course books. The words are such as accent, adjective, adverb, conjunction, dictate, drill, fluency, modal, prepositions, pronunciation, punctuation, syllable, and vocabulary. Secondly, the non-high-frequency English words appear in the course books because they are specific words that are connected with the theme in a unit of the books. For example, when the topics of Chapter 4 of the course book for grade seven are things in a classroom, the non-high frequency words that appear in the course books are such as notebook, rack, sharpener; in a kitchen saucepan, shallots, spatula, stool, tablespoon, around the house and the yard, bathroom, bathtub, barrow, hoe, etc. Thirdly, the authors of the books select low-frequency words instead of high-frequency ones, for example, adorable, awesome, and burrows. These words can be replaced by high-frequency ones namely attractive, excellent, and holes respectively. Fourthly, the words that now belong to high-frequency words were not included in the high-frequency ones of GSL and Academic Word List. The evidence is that 85 of 625 words not in the lists belong to the words in the NGSL. There are 2,809 word families in the NGSL (Browne, 2013). 1,496 of the NGSL are word families in the GSL and 437 word families in Coxhead’s academic word list.

The last finding of this study shows that more than half of the words in the three course books only appear in one of the three course books. This means that the users of the course books will not meet the words in the other two course books. In addition to that, the frequency of word occurrence in each of the course books is also dominated by one time except for the first thousand most frequent words, where more than half of these words appear five times or more in the course books for grades seven to nine. This is not surprising since the first thousand most frequent English words, which belong to the General Service List and the New General Service List, appear in any English text and its coverage in English texts is around 85% (Coxhead, 2000) and around 90% (Browne, 2013) respectively. The ten most frequent words of the first thousand words are functional words, such as articles a, an, preposition of, in, the verb ‘be’, pronouns
it, they, etc. This fact is in line with the result of the study on three types of texts in senior high school course books: descriptive, recount, and narrative conducted by Rustan and Andriyanti (2021).

The percentages of the words belonging to the second thousand most frequent words, academic words, and not in the lists that appear five times or more are much lower than that of the first thousand most frequent ones. Conversely, the percentages of the words that appear less than five times and even once in these groups are much bigger. This is likely because the topics of the course books are different from one book to another book, from one unit to another unit.

A newly-known word will be remembered if it is encountered at least seven times over an interval of time (Gaybulloyevna, 2022) and 10 times (Web, 2007). Web found that the greater the repetition the fuller the vocabulary knowledge students will get. Schmitt et al. (2017) state eight to ten times seems to be the minimum number for consistent incidental learning. Thus, it is probable that most of the English words in the course books will not be fully learned and acquired by the users of the books if there are no other efforts made by the students and teachers to make the students understand the words well. Nation (2015) argues that to know words, including their forms, meanings, and uses, learners need to meet the words many times in different contexts, and give intense attention to the words. For example, they need to find out how the words are pronounced, what their meanings are, how they are used, and how they are retrieved.

CONCLUSION AND SUGGESTION

The total number of English word families introduced in the English course books for JHS students is 1,988. It is sufficient for reading simplified picture stories, but not for reading and listening comprehension of authentic written and spoken texts. Based on the English words in the course books, the teacher and the students are demanded to develop the words through challenging activities so that the number of the word types and word families of the students increases significantly to reach the threshold.

The English words in the course books belonging to the first and second-thousand most frequent words and academic words do not reach the adequate numbers, that is, around 80% of those words. This suggests that the teacher and students have to make efforts to increase students’ vocabulary knowledge by using any possible teaching media and activities. In addition, future authors of English course books for JHS students should include the first and second thousand most frequent words and academic words adequately. Conversely, as the number of English words that do not belong to the three lists exceeds those in authentic texts, the teacher and students need to analyze which of those words
require elaboration. The teachers using the books should know the GSL and NGSL so that they can direct their teaching to the mastery of English words in the lists.

The frequency of most English word occurrences in the course books is unlikely to lead to incidental word learning. Most of the English words in the three English course books appear only in one of the course books and most of the English words have a frequency of appearance in each of the course books less than five times. It is necessary for English teachers to create classroom activities, such as mind mapping, brainstorming of the low-repetition words connected with the forms, meanings, and use of the words to increase the repetition of the words. The authors of future English course books should also consider the frequency of word occurrence when including English words in the course books and provide a number of activities that demand the students to process and use the words repeatedly in written and spoken forms.

REFERENCES


Copyright © Year by Author, e-ISSN 2657-1617


Syarofi, A., & Shobaha, S. (2023). The implementation of academic word list and its implication to the improvement of EFL students’ academic writing
quality. Lingua Cultura, 17(1), 17-23. https://doi.org/10.21512/lc.v17i1.9022


Authors’ Brief CV

Ari Nurweni is a lecturer at the English Education Department of Universitas Lampung. She completed her bachelor's in English Teaching at Universitas Negeri Yogyakarta, Post-graduate Dip TESL and Master’s in applied linguistics at Victoria University of Wellington, and Doctoral degree in linguistics at Gadjah Mada University. Her research interest is in vocabulary.

Endang Komariah is a lecturer at the English Education Department of Universitas Lampung. She completed her bachelor’s in English Teaching at Universitas Lampung, her master’s in English Language Teaching at Malang University, and short courses in English syllabus designs and language assessment at RELC SEAMEO. Her research interests are in TEFL and language assessment.