Textbook analysis using Vocabulary Profiler

Gusti Astika
Universitas Kristen Satya Wacana
Salatiga, Indonesia
Abstract
This presentation describes the uses of the Vocabulary Profiler to profile the vocabulary used in a textbook currently used in Indonesian Junior High School. The sample text book is *English in Focus* for grade VII. The text book has been recommended by the Ministry of Education and Culture for use in Indonesian Junior High Schools. The overall profile shows that vocabulary coverage is 88.20%, for the first (K-1) and second (K-2) most frequently used words indicating that it is below the necessary level (95%) for easy comprehension. The analysis also reveals negative vocabulary profile, that is proportions of word families in K1 and K2 categories that are not found in the text books.

**Keywords:** vocabulary, word frequency, vocabulary profiler, negative vocabulary.
Textbook analysis using Vocabulary Profiler

Introduction

In the first or second language setting, learners use English to communicate and they learn vocabulary based on their needs and motivation. In an EFL context, vocabulary learning is learned as a compulsory subject at schools and learning is constraint by time and school context. Language learning is structured that demands specific approaches, methods, and techniques. EFL learners may not need to know and use all vocabulary in the text book. The goal of learning has been predetermined in the school curriculum and learners will only learn and use a limited number of vocabulary covered in the text book. The question that teachers will ask is which vocabulary items are actually needed and how to determine and select those vocabulary items. Students are often frustrated when they are confronted with a large amount of new vocabulary in the textbook. The teachers’ task in teaching vocabulary seems equal to the learning burden of the students when they have to decide which words to introduce, how to select words that are actually needed by the students, and when during instruction new words are appropriately introduced. Moreover, when other things are considered such as the amount of effort, time, and money both the learners and teachers have committed to learning, it is imperative that a principled instructional vocabulary learning has to be designed to make vocabulary learning more effective and enjoyable. The importance of teaching vocabulary has been widely recognized. According Nation (2001) acquisition of 2000 high-frequency words should be sufficient for comprehension of texts in English text books.
Word Frequency

Nation (1990:4) describes that vocabulary can be divided into high-frequency words, low-frequency words, and specialized vocabulary. Based on this frequency level, teachers need to decide which words are needed by the learners, how to select words for teaching, and how often those words should be exposed to the students for acquisition. Some studies (Stæhr, 2008; Coxhead, 2011; Schmitt, 2000; Nation, 2001 and Horst, 2013) have suggested that at the early years of learning high priority should be placed on high-frequency words because this is an important learning goal and the teaching of these words has to be explicit. Vocabulary knowledge forms a fundamental basis for proficiency in second or foreign language learning. Lacking in vocabulary knowledge will hinder students from communication in the second or foreign language because vocabulary is a necessary component for improving all areas of communication (Godwin-Jones, 2010).

Word frequency can conveniently be identified using the Vocabulary Profiler (www.lextutor.ca). The profiles can provide information about the vocabulary items that belong to K-1 words; the most frequently 1,000 words, K-2 words; the next most frequently 1,000 words, K-3, the third most frequently used words, AWL; the Academic Words List, and Off-list words; the words that do not belong to the classifications. The K-1 words are the most common and easiest words found in any text. The K-2 and K-3 words are those words that are less common and less frequent in a text; therefore, these K-2 and K-3 words are more difficult than K-1 words. The AWL words are those words that are commonly used in academic texts. The Off-list words are usually names of places or people or specialized terms used in a specific discipline. Each of these classifications can also be used to provide information about the
proportion (percentage) of the words used in a textbook in comparison to the well-established word lists, the New General Service List (http://www.newgeneralservicelist.org)

Profiling the Vocabulary

The sample text book for analysis is *English in Focus* book VII, one of the English textbooks for Junior High Schools that have been documented in the Ministry of Education website (http://bse.kemdikbud.go.id). The text book is in a pdf file format, therefore, it is necessary that it has to be converted into a doc. file format. The proper names, numbers, phonetic symbols and the preface page that use Indonesian language are excluded for analysis before it is fed into the Vocabulary Profiler.

Below are the procedure for using the Vocabulary Profiler and the screen shots outputs of the analysis.

1. Visit the website of the Vocabulary Profiler at www.lextutor.ca to see the first screen.

2. Screen 1. Vocabulary Profiler page
Please note that there are several tools that can be used to learn and analyze vocabulary.

For the purpose of this discussion, the tool that is needed is ‘Vocabprofile’ in the second column of this screen page.

3. The next step is to hit ‘Vocabprofile’ to get the following page

Screen 2. Vocab Profile Home.

There are three profilers on this page: VP Classic, VP-kids, and VP-Compleat. Since the textbook sample used in this analysis is for students at Junior High School, VP-compleat is selected instead of VP-Classic because the VP-compleat is the current development version of the VP-Classic that can produce more precise classification of word frequency.
4. Choose ‘VP-Compleat’ to show the instruction to feed the text into the profiler as can be seen in Screen 3 below.

![Screen 3. Web page of VP-Compleat](image)

The instruction to use the profiler can be seen in the white space of the screen. As mentioned earlier, the text that has been edited from the text book can now be pasted in the white part of the screen after deleting the instruction. After selecting the ‘Submit Window’ button, the output of the analysis will look like the following screen. This is the vocabulary profile of the text book, *English in Focus*, mentioned above.
The first row in Screen 4 shows three terms: lemmas, types and tokens. Lemma is head word, for example: the head word of discussion and discussing is discuss. Type is different words, for example: committee and comparison are different words. While discussion and discussing, or discussed are considered as the same type. Token is words in a text; the total number of words in a text. For example, if in a text there are committee [2], discussion [1], really [3], and students [5], the number of token is 11. The first column shows the different frequency levels: NGSL_1 (New General Service List 1) or K-1, is the word group that is most frequently used. NGSL_2 or K-2, is the second most frequently used words group. NGSL_3 or K-3, is the third most frequently used words group. NAWL (New Academic Word List) is words group that is commonly used in academic texts. Off-list is words that do not belong to any of the groups.
The statistics in Screen 4 shows that more than two-third (81%) of the vocabulary used in the textbook fall within the most frequently used 1000 words group (K-1). With the additional K-2 word coverage (7.20%) the cumulative percentage of the word coverage is 88.20%, which is below the desired level for good comprehension of the texts in the book. According to Hirsch (2003), an understanding of 95% of the words is necessary for comprehension. Based on the data above, good comprehension of the text book should include knowledge of words that belong to K-3 group as much as 3.74%, knowledge of words that belong to academic words as much as 0.98% and some words in the Off-list group. We need to question whether the academic words as many as 203 words used in the text book need to be introduced to the students in Junior High Schools. Another point that is worth considering is the number of off-list words that reaches 1459 words. Although this off-list category is excluded from the frequency list and may occur infrequently (low frequency words), it may have words that students at this level need to know. This low frequency list should not be ignored in teaching and teachers have to make a selection for useful words in this category.

Screen 5, 6, and 7 below show the words that have been automatically grouped and color-coded according to their frequency levels. As seen in the output, the words have been alphabetically organized each with its frequency of occurrences in square brackets. The total number of words in NGSL-1 (K-1), for example, is 16721, or 81% of the total number of words in the text book. Of this number, there are 1036 types and 687 lemmas (head word).
Negative vocabulary

Besides word frequency, the Vocabulary Profiler can also be used to identify negative vocabulary, that is the vocabulary in the New General Service List (NGSL) (http://www.newgeneralservicelist.org), that is not found in the text book. Below are the screenshots of negative vocabulary in the K-1, K-2, and K-3 frequency levels. Each of the Negative Vocabulary Profile shows the summary statistics. In Screen 8, for example, the total number of word families (lemmas) in K-1 is 1001 words, the number of word input (words in the text book) in K-1 category is 687 (68.63%), and the number of K-1 words not found in the text book is 315 (31.47%). The words in ‘Found’ and ‘Not Found’ columns can be extracted by clicking the header to produce the list of all words under each column.

As can be seen in the negative vocabulary tables, the percentage of negative vocabulary, that is the vocabulary not found in the textbooks based on the New General Service List, increases from K-1 to K-2 and K-3 frequency groups. The reason for this increase is the decrease of vocabulary coverage in the textbooks from K-1 to K-2 and K-3. In other words, the number of
high frequency vocabulary (K-1) is larger than that of the second and third most frequently used words (K-2 and K-3). It appears that this is to be expected because at the Junior High School level, the priority of vocabulary learning should be those words in the K-1 group, while the words in K-2 and K-3 groups may be postponed until the acquisition of the K-1 words.
Screen 10. Negative vocabulary of K-3

The information provided in these tables may point to the need for an assessment of the textbook which takes into account the number of negative vocabulary (351 word families or 31.47%) in the K-1 group. As indicated in the tables, the word families not found in the textbook is close to three-fourth (70.27%) of the number of word families included in the NGSL. A selection has to be made to determine which vocabulary items are actually needed and relevant at this level of education (Junior High School). The vocabulary not found under K-3 frequency group is even higher, reaching 81.75% or 654 words. In my view, the negative vocabulary items in K-1, K-2, and K-3 have to become the attention of the teachers or book writers in order to enlarge the students’ vocabulary size.

Comparing chapters in the text book

The third tool in the Vocabulary Profiler that is also very useful for vocabulary analysis is ‘Text Lex Compare’ available in the third column of the Vocabulary Profiler website. This tool can be used to compare the vocabulary in the book chapters. The comparison will show the token recycling index of the chapters being compared. Recycling index is the ratio between words that
are shared by two chapters and the total number of words in the second chapter. This index provides useful information about what words are similar or shared in both chapters and what words are new or unique in the second chapter. For teaching purposes, teachers need to pay attention to those words that are unique in the second chapter.

The comparison below is between chapter 3 and chapter 4 in the text book, *English in Focus*. The analysis of comparison shows that the token recycling index is 77.77%, indicating as much as 77.77% of words in chapter 3 and chapter 4 are similar (shared). Thus, new or unique words in chapter 4 is 22.23% (100-77.77%) or 764 words. The comparison output is displayed below.

The figure next to a word is the occurrences or frequency of that word in the chapter.

<table>
<thead>
<tr>
<th>Unique to first</th>
<th>Shared</th>
<th>Unique to second</th>
</tr>
</thead>
<tbody>
<tr>
<td>372 tokens</td>
<td>2673 tokens</td>
<td>764 tokens</td>
</tr>
<tr>
<td>196 families</td>
<td>305 families</td>
<td>386 families</td>
</tr>
<tr>
<td>001. prohibit 14</td>
<td>001. the 205</td>
<td>Freq first</td>
</tr>
<tr>
<td>002. late 11</td>
<td>002. be 184</td>
<td>(then alpha)</td>
</tr>
<tr>
<td>003. drive 9</td>
<td>003. you 135</td>
<td>001. sell 17</td>
</tr>
<tr>
<td>004. message 9</td>
<td>004. a 104</td>
<td>002. polite 13</td>
</tr>
<tr>
<td>005. dream 8</td>
<td>005. i 88</td>
<td>003. one 12</td>
</tr>
<tr>
<td>006. nine 7</td>
<td>006. to 82</td>
<td>004. shirt 11</td>
</tr>
<tr>
<td>007. police 7</td>
<td>007. in 60</td>
<td>005. assist 9</td>
</tr>
<tr>
<td>008. announce 6</td>
<td>008. of 59</td>
<td>006. pardon 9</td>
</tr>
<tr>
<td>009. fax 5</td>
<td>009. this 47</td>
<td>007. apple 8</td>
</tr>
<tr>
<td>010. month 5</td>
<td>010. do 38</td>
<td>008. blue 8</td>
</tr>
<tr>
<td>011. play 5</td>
<td>011. practise 37</td>
<td>009. it’ 8</td>
</tr>
</tbody>
</table>

**Conclusion**

Vocabulary profiler could play a significant role to help teachers to develop learners’ vocabulary knowledge in a way that is different from more conventional way of teaching
vocabulary where teachers may rely on their teaching experiences and intuitive knowledge. This expert judgment in teaching vocabulary may serve as a short-cut in teaching. However, words selected in this way may or may not be those words that the students need to acquire at this level. The cumulative proportions of vocabulary frequency groups serves as a handy tool to determine the relative difficulty of a textbook because the output of the profiler indicates the vocabulary coverage of the textbook.

The negative vocabulary points to the needs for more coverage of vocabulary items that learners should acquire, at least those words in the K-1 group. These ‘missing’ words should be taught to the students earlier than the other word groups (K-2, K-2, and K-3) since they are essential for text comprehension. The inclusion of these words in the teaching materials may require teachers’ creativity and expert professional decision during teaching and learning process within the time frame available for the whole program.

This vocabulary analysis has been limited only to one book. A similar research needs to be conducted with more samples of textbooks currently used. More studies such as this one would provide teachers and researchers in this area with more data and findings for the betterment of teaching and learning English vocabulary in our schools.

REFERENCES


Vocabulary Profiler. [www.lextutor.ac/vp](http://www.lextutor.ac/vp).