

WHAT WORDS SHOULD WE TEACH? EXPLORATION INTO VOCABULARY PROFILER

Universitas Kristen Satya Wacana
Salatiga
gustiaastika@yahoo.com

Abstract

Students learning English as a foreign language are often frustrated when they have to cope with new words. The teacher's task in teaching vocabulary seems equal to the student's learning burden when it comes to decide which words to introduce and how to select words that the students actually need to learn. This article proposes that *Vocabulary Profiler* (www.lexxtutor.ca) can be utilized to profile the vocabulary in a textbook to produce word frequency. An English textbook with 21,577 words was selected for analysis with the Vocabulary Profiler. The output shows high and low frequency word groups that can be used as bases for vocabulary selection in teaching. In addition, the output provides information about negative vocabulary and token recycling index, an indicator of text comprehensibility. Although this article sampled a textbook for use in Indonesian contexts, the ideas might be of interest to EFL teachers in other countries.

Keywords: vocabulary, word frequency, vocabulary profiler, negative vocabulary.

DOI:[10.30957/ijotl.v2n3.386](https://doi.org/10.30957/ijotl.v2n3.386).

1. INTRODUCTION

In the first or second language setting, learners use English to communicate and they learn vocabulary based on their 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 textbook. 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 textbook.

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 school textbooks.

Nation (1990:4) classifies vocabulary into high-frequency words, low-frequency words, and specialized vocabulary. When teaching, 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. Word frequency can be identified using the Vocabulary Profiler (www.lex tutor.ca). The profile provides 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. K-1 words are the most common and easiest words found in any text. 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. AWL words are those words that are commonly used in academic texts. 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 words used in a textbook in comparison to the well-established word lists, the New General Service List (<http://www.newgeneralservicelist.org>)

2. REVIEW OF LITERATURE

Based on the frequency level proposed by Nation (1990), teachers need to decide which words are important to the learners, how to select words for teaching, and how often those words should be exposed to the students for acquisition. The decision for teaching will determine the pedagogical activities to achieve the learning goals. Some studies have recommended that high-frequency words should be the priority in teaching and taught early especially at the beginning years of learning (Coxhead, 2011; Douglas, 2013; Stæhr, 2008, and Horst, 2013). High-frequency words are very essential in language learning, therefore, instruction has to focus on the high-frequency group, then continue to words at a lower frequency category. These studies show that teaching beginning learners should focus on 2000 vocabulary level (high frequency group) and explicit teaching is needed. In light of this principle, vocabulary teaching has to be based on frequency-informed decisions. Such a decision can be implemented by utilizing Lextutor (www.lex tutor.ca).

Teaching vocabulary in our schools at present may not be based on frequency-informed decisions. English teachers may not be aware that high frequency words are essential and it does not need systematic teaching. Teachers may think that students will acquire them incidentally during reading or speaking activities. While incidental learning may occur during learning, researchers such as Horst, Cobb & Meara (1998), as cited in Horst (2013), argue that this incidental teaching of vocabulary is not sufficient and the speed of students' vocabulary knowledge is slow. This argument is even more evidenced in EFL contexts where classroom exposure for vocabulary acquisition is limited. In fact, according to Schmitt (2000), Nation (2001), and Horst

(2013), high-frequency words have to be taught quickly because it is worth-it and pedagogically essential.

With the increasing needs for English in education in general and its strong position in the school curriculum, the importance of vocabulary in language learning should receive more recognition. Research in this area has come out from corpus studies (Horst & Nicolae, 2013) showing the need for learners to know most frequent words in the language being learned. Similar studies (Zhang & Li, 2011; Read, 2004) have also revealed that vocabulary knowledge forms a fundamental basis for proficiency in second 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).

In order to comprehend a text learners need to know about 95-98% of the words (Hirsch, 2003; Laufer & Ravenhorst-Kalovski, 2010). Research on vocabulary acquisition has revealed that there is a high correlation between word knowledge and reading comprehension (Sedita, 2005; Schmitt's and Grabe, 2011). For learners in an EFL environment such as that in Indonesian schools, the only source of input for vocabulary development is, to a large extent, from school textbooks for receptive vocabulary knowledge. Schmitt (2000) suggests that at early stages of learning, students should learn about 1000-2000 high-frequency words and then increase to about 3000-5000 word families necessary to read authentic texts that may include academic words for reading texts at university level. It needs to note that learners have more difficulty with developing productive vocabulary than receptive vocabulary (Zheng, 2012). Adolphs and Schmitt (2004) estimates that learners should at least master 2000 word forms in order to be able to understand spoken discourse in different contexts. As Schmitt (2000) acknowledges, vocabulary acquisition is a developmental process especially noticeable in an acquisition-poor environment. Such an environment can not provide learners with adequate opportunity to use English words they learn in class. This condition, in turn, may weaken their motivation to develop communicative skills because their language learning outside class is not socially functional; hence, their vocabulary development may be diminishing.

The discussion above points to the needs for a principled and systematic analysis of vocabulary that can be effectively used for the development of students' vocabulary knowledge. Therefore, this study used the Vocabulary Profiler (www.lex tutor.ca) that has been widely applied to determine the profiles of the words in an English textbook currently used in junior high schools. The profile will provide useful information to teachers to select vocabulary items that are relevant for learning in a principled way. With the background discussed above, this study had the following questions,

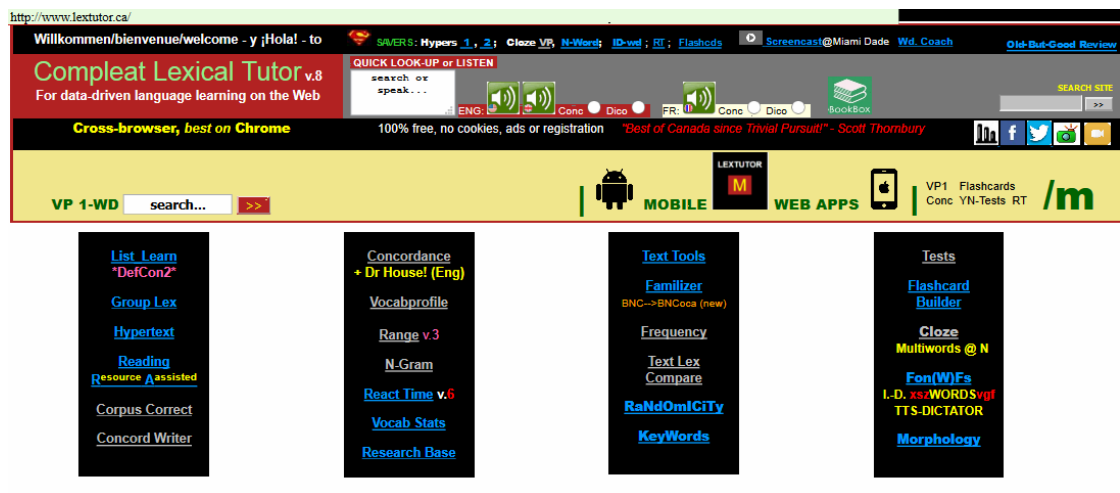
1. What is the profile of the vocabulary used in the textbook?
2. What is the proportion of the negative vocabulary in the textbook?
3. What is the token recycling index of two chapters in the textbook if they are compared?

3. METHODS

This study aimed to identify the vocabulary items that have to be prioritized in teaching by classifying the words in the coursebook sample into frequency levels using *Vocabulary Profiler*, and to identify the negative vocabulary items in the *New General Service List* (<http://www.newgeneralservicelist.org>), that were not found in the course book. The data for the study were the vocabulary items in the course book with the title, *English in Focus* for grade VII of Junior High School in Indonesia. The proper names, numbers, phonetic symbols and the preface in Indonesian were excluded from the analysis because they were not relevant to the frequency counts. The vocabulary items were typed into a word-file format and then copied into the *Vocabulary Profiler* for analysis.

The following sections present the procedure for using the *Vocabulary Profiler* and the analysis of the vocabulary items in the course book. 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.lex tutor.ca to see the first screen.

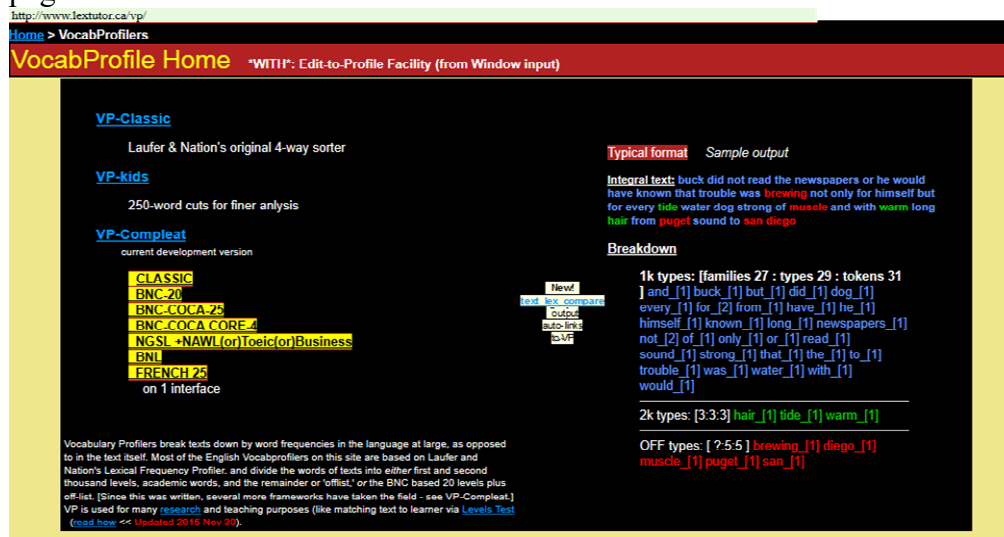


Screen 1. Vocabulary Profiler page

Please note that there are several tools that can be used to learn and analyze vocabulary. To start the analysis, 'Vocabprofile' in the second column of this screen page is selected.

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

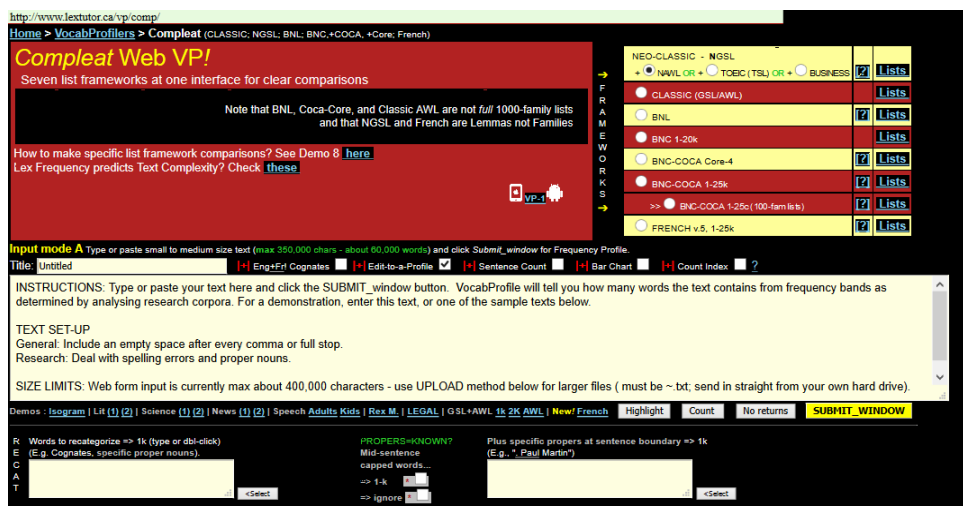
page



Screen 2. Vocabulary Profile Home

There are three profilers on this page: VP Classic, VP-kids, and VP-Compleat. Since the course book sample used in this analysis is for students at Junior High School, select ‘VP-compleat’ instead of VP-Classic because it is the current development version of the VP-Classic. It can produce more precise classification of word frequency.

3. Choose 'VP-Compleat' to see the next page as can be seen in Screen 3 below.



Screen 3. Web page of VP-Compleat

The instruction to use the profiler can be seen in the white space of the screen. The edited course book can now be copied into the blank section after deleting the instruction.

4. FINDINGS AND DISCUSSION

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 a text has *committee* [2], *discussion* [1], *really* [3], and *students* [5], the number of token is 11. The first column the 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 three-fourth (81%) of the vocabulary used in the course book 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 textbook should include knowledge of words in K-3 group as much as 3.74%, knowledge of words in AWL group 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,

used in the course book need to be introduced to the students at 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 4 below is the vocabulary profile of the course book.

<http://www.lex Tutor.ca/cgi-bin/vp/comp/output.pl>

Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas]	687 (55.67)	1036 (45.34)	16721 (81.00)	81.00
NGSL_2 [1000 lemmas]	298 (24.15)	366 (16.02)	1486 (7.20)	88.20
NGSL_3 [801 lemmas]	182 (14.75)	213 (9.32)	773 (3.74)	91.94
NAWL [963 lemmas]	67 (5.43)	75 (3.28)	203 (0.98)	92.92
Off-List:	??	595 (26.04)	1459 (7.07)	99.99
Total (unrounded)	1234+?	2285 (100)	20642 (100)	≈100.00

Screen 4. Vocabulary profile output

Below are a small proportion of word samples in each frequency group derived from the statistics in the Vocabulary profile output (Screen 4) above. The vocabulary in each group has been arranged alphabetically with the number of occurrences in the square brackets. A complete word list of word frequency groups can be found by scrolling down the page of the vocabulary output in the online version of the profile.

NGSL-1 or K-1 [lemmas 687 : types 1036 : tokens 16721]

a [62] able [34] about [79] above [2] accept [3] accord [1] achieve [2] across [1]
act [14] action [7] activity [15] actually [1] add [14] address [10] advantage [1]
after [61] afternoon [19] again [4] against [1] age [2] agency [2] ahead [1] all [49]
allow [1] almost [1] along [1] already [1] also [16] always [9] among [1]
amount [1] and [312] animal [2] another [5] answer [75] any [11] anyone [1]
anything [2] anyway [3] apply [1] ...

NGSL-2 or K-2 [lemmas 298 : types 366 : tokens 1486]

accommodation_[1] accompany_[2] ad_[5] additional_[1] advertisement_[23]
advise_[1] afraid_[4] amaze_[1] ancient_[2] annual_[1] appointment_[2]
appropriate_[5] arrange_[6] artist_[1] assistant_[16] attach_[3] attitude_[2] attract_[3]
audience_[1] bag_[9] ball_[9] beach_[5] beauty_[1] bedroom_[20] bike_[3] bird_[3]
birth_[2] blood_[2] borrow_[4] bottle_[5] bottom_[1] brand_[1] breakfast_[7]
breath_[2] bright_[1] broad_[1] brown_[6] busy_[3] ...

NGSL-3 or K-3 [lemmas 182 : types 213 : tokens 773]

accomplish_[2] airline_[2] alright_[1] announcement_[6] anymore_[1] apologize_[3]
architecture_[1] athlete_[2] august_[1] aunt_[7] badly_[2] bath_[3] beside_[4]
billion_[2] bin_[1] bowl_[6] bread_[5] broadcast_[1] bunch_[1] buyer_[1] cake_[9]
chase_[1] cheek_[3] cheese_[5] chicken_[10] chocolate_[7] cigarette_[2] clause_[1]
clothing_[1] comedy_[1] compose_[1] consult_[10] consume_[2] cough_[1] cousin_[4]
cow_[1] crack_[1] cream_[6] ...

NAWL or Academic Word [lemmas 67 : types 75 : tokens 203]

absorb_[1] activate_[1] apple_[10] bargain_[3] basin_[1] beam_[1] blank_[17]
bracket_[2] cheat_[3] chess_[2] cinema_[1] classroom_[8] clay_[1] click_[1] clue_[4]
composite_[1] cone_[1] consonant_[2] correctly_[12] drain_[1] enzyme_[1] fabric_[4]
goods_[1] gram_[1] harvest_[2] homework_[2] horizon_[4] identification_[2]
informal_[12] insect_[4] intensity_[1] junior_[2] leaf_[4] leisure_[2] lung_[2] mall_[1]
mechanic_[2] neat_[2] neutral_[2] nicely_[1]...

OFF-LIST: [?: types 595 : tokens 1459]

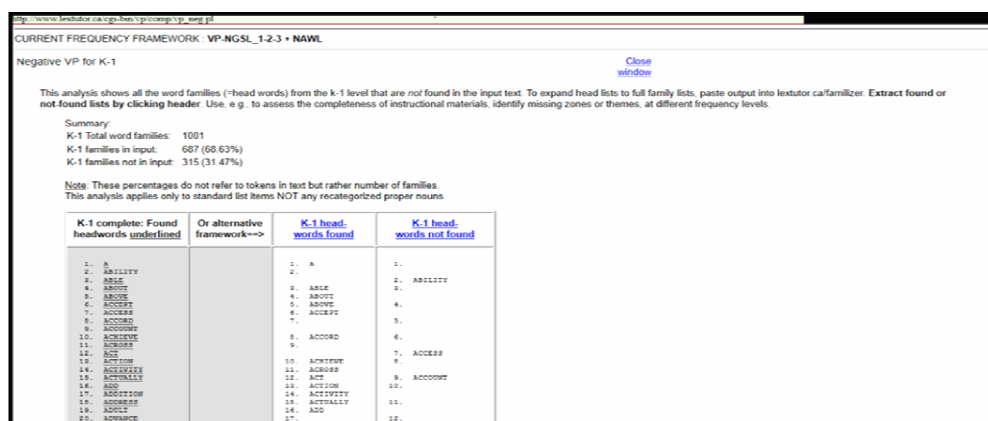
aboard_[1] accelerator_[3] excellence_[1] accessories_[2] accidentally_[2]
adjectives_[1] adrenaline_[1] adverb_[5] adverbs_[3] aero_[1] airways_[2] alike_[1]
allergy_[1] aloud_[28] alphabets_[1] american_[1] announcer_[1] antenna_[1]
apology_[13] appliances_[1] appropriately_[3] asleep_[3] atlantis_[1] attendant_[2]
automobiles_[1] awake_[2] axles_[1] backpack_[2] badge_[2] bake_[2] baker_[1]
bakery_[1] baking_[1] banana_[1] barber_[2] barn_[1] barns_[1] baseball_[3] ...

Referring to Adolphs and Schmitt's (2004) suggestion, those words in NGSL-1 (K-1) and NGSL-2 (K-2) should be prioritized and taught early because they are necessary for understanding of spoken discourse, while those words in NGSL-3 (K-3) and AWL can be introduced when learners need to read and understand authentic text (Schmitt, 2000). Important words for teaching can also be selected from the Off-List group. This is the 'prerogative' of the teacher to determine what words are relevant and needed by the students at Junior High School level. For example, the word *aloud*, that is often used in giving instructions, or *barber*, that is often seen in a large artistic display on a barbershop window. Still, another useful word is *chocolate*, that students can

remember easily. So, selection can be made on the bases of the needs for instruction, students' life experience, survival, or immediate needs.

4.1. Negative vocabulary

The second question this study attempted to answer is the number of negative vocabulary. Negative vocabulary is the vocabulary in the New General Service List (NGSL) (<http://www.newgeneralservicelist.org>), that is not found in the course book. Below is the screenshots of negative vocabulary in the K-1 frequency level.



Screen 5. Negative vocabulary of K-1

The information in Screen 5 shows the total number of word families (lemmas) in K-1, which is 1001 words, the number of word input (word family in the textbook) in K-1 category, 687 (68.63%), and the number of K-1 words not found in the textbook is 314 words (31.36%). 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. The complete information of negative vocabulary of K-1, K-2, K-3, and AWL groups is presented in Table 1 below.

Table 1. Negative vocabulary

	K-1 words	K-2 words	K-3 words	AWL
Total head words	1001	1000	801	963
Total head words in text	687 (68.63%)	298 (29.80%)	147 (18.35%)	67 (6.95%)
Negative vocabulary	314 (31.37%)	702 (70.20%)	654 (81.65%)	896 (93.05%)

As can be seen in the negative vocabulary table, the proportion of negative vocabulary 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 textbook 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.

The information provided in Table 1 may point to the need for an assessment or revision of the textbook which takes into account the high frequency words, for example, 314 negative vocabulary (31.37%) in the K-1 group. As indicated in the table, K-2 word families not found in the textbook is more than two-third (70.20%) the number of K-2 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 negative vocabulary in K-3 frequency group is even higher, reaching 81.65% or 654 words. In my view, the negative vocabulary items of the high frequency groups in K-1, K-2, and K-3 have to become the attention of teachers or book writers and consider to include them in the textbook in order to enlarge the students' vocabulary size.

4.2. Comparing chapters in the course book

The chapters or units in a course book are usually presented in a chronological order. When it comes to introducing new words, the selection of new vocabulary in the succeeding chapter can become a tedious work. The tool in the Vocabulary Profiler that is very useful for this purpose is *Text Lex Compare* available in the third column of the Vocabulary Profiler website, as seen below.

Screen 6. Text Lex Compare

This tool can be used to compare the vocabulary in two chapters of a course book. The texts to be compared are copied into the clean white parts of the page, then hit the 'Submit' button. 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.

Table 2 below shows a portion of the comparison between chapter 3 and chapter 4 in the sample course book, *English in Focus*. The screenshot of the comparison output is displayed below.

Table 2. The output of text comparison

TOKEN Recycling Index: (2673 repeated tokens : 3437 tokens in new text) = 77.77%

FAMILIES Recycling Index: (305 repeated families : 691 families in new text) = 44.14%

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

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). This index is also a measure of text comprehensibility. Thus, new or unique words in chapter 4 is 22.23% (100%-77.77%) or 764 words. The figure next to a word is the occurrences or frequency of that word in the chapter. For teaching chapter 4, teachers need to select those words in the third column that are unique or new. This table can be very useful when teachers decide to introduce new vocabulary. The priority may be given to those words with high frequency of occurrences, then go on to lower frequency words.

5. CONCLUSION

Vocabulary profiler could ease teachers' work in helping learners to develop vocabulary knowledge in a way that is different from more conventional way of teaching vocabulary in which teachers may rely on their teaching experiences and intuitive knowledge. This expert judgment in teaching vocabulary may serve as a short-cut. 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 reference to determine the relative difficulty of the textbook because the output of the profiler indicates the vocabulary coverage in 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, and K-3) since they are essential for text comprehension. The inclusion of these words in 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 course 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

- Adolphs, S. and Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards and B. Laufer (Eds.), *Vocabulary in a Second Language* (pp.39-52). Amsterdam: John Benjamins.
- Buku Sekolah Elektronik. <http://bse.kemdikbud.go.id>.
- Coxhead, A. (2011). The academic word list 10 years on: research and teaching implications. *TESOL Quarterly* 45(2), 355-362.
- Douglas, S. R. (2013). The lexical breadth of undergraduate novice level writing competency. *The Canadian Journal of Applied Linguistics*, 16 (1), 152-170.
- Godwin-Jones, R. (2010). Emerging technologies, from memory places to spacing algorithms: approaches to second language vocabulary learning. *Language Learning and Technology*, 14, 4-11.
- Hirsch, E.D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, Spring. American Federation of Teachers.
- Horst, M. (2013). Mainstreaming second language vocabulary acquisition. *The Canadian Journal of Applied Linguistics*, 16 (1), 171-188.
- Laufer, B., and Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15-30.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers..

- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- New General Service List. [Http://www.newgeneralservicelist.org](http://www.newgeneralservicelist.org).
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- Schmitt, N. (2000) *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2004). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., Jiang, X. and Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95, 26-43.
- Sedita, J. (2005). Effective vocabulary instruction. *Insights on Learning Disability*, 2(1), 33-45.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36, 139-152.
- Vocabulary Profiler. www.lex tutor.ac/vp.
- Zheng, Y. (2012). Exploring long-term productive vocabulary development in an EFL context: the role of motivation. *System*, 40, 104-119.
- Zhang, B. and Li, C. (2011). Classification of L2 vocabulary learning strategies: evidence from exploratory and confirmatory factor analysis. *RELC Journal*, 42, 141-154.