THE ENGLISH VOCABULARY SIZE AND LEVEL OF ENGLISH DEPARTMENT STUDENTS AT KUTAI KARTANEGERA UNIVERSITY

Sudarman
Chulalongkorn University; Lembaga Pengelola Dana Pendidikan, Indonesia
sudarmannasir@gmail.com

Sumalee Chinokul
Chulalongkorn University
csumalee@chula.ac.th

Abstract The study employed descriptive-quantitative research design which aimed to investigate the mean vocabulary size and mastery level of English department students. The subjects of the study were the first-year students majoring in English education department at Kutai Kartanegara University. Thirty-three students employed in this study were given the Vocabulary Levels Test (VLT) to measure their English vocabulary size as well as their mastery level. The findings of this study revealed that the students only knew about 1,273 word families. The result was still below the threshold as suggested by the vocabulary scholars and was not improved after years of learning. The finding also showed that the participants had a very low mastery level. They did not even master the 2,000 or 3,000 high frequency word level, nor the academic vocabulary level. The results indicate that the participants had a low vocabulary proficiency and suggest the future research to focus on investigating the vocabulary learning and instruction strategies that are effective to develop the students’ high frequency words and academic words.

Keywords: vocabulary size, vocabulary level, English department students, vocabulary proficiency

A. Introduction

The growing importance of English as one of international languages as well as a lingua franca has affected educational curriculum of all countries in the world. In specific case of Indonesia, the growing importance of English can be seen from the increasing number of schools, from kindergarten to university level, using English as a medium of instruction (Dardjowidjojo, 2002). In addition, English is officially taught as a foreign language in junior high schools for 3 years and in senior high schools for 3 years. Despite not being a compulsory subject, English is still taught in primary schools as the local content subject (Sikki, Rahman, Hamra, & Noni, 2013). Although it has been taught for more than 6 years, Indonesian students’ proficiency in English remains very low (Kirkpatrick, 2012). The English Proficiency Index (EPI) data arranged by English First (EF) show that in 2017 Indonesia ranked 39th out of 80 countries in the world and 10th out of 20 countries in Asia. The Indonesian score was 52.15 on average and falls under ‘low proficiency band’ level.
Saukah (2016) reveals that Indonesian English department students also have low proficiency. It is quite shocking because as English teachers to be, they are supposed to be proficient in general English and classroom English (Richards, 2017). According to Renandya (2018), a professional language teacher should at least have sufficient content knowledge, pedagogical content knowledge, and sufficient proficiency in the language to be able to teach through the target language effectively. Although language proficiency does not automatically guarantee effective teaching, research shows that teachers with a high level of proficiency seem to be more capable at using the language to perform the classroom tasks and provide an appropriate language support to their students.

Ivone (2005) assumes that the lack of vocabulary knowledge could be one of factors that contributes to the low level of proficiency gained by Indonesian students. This factor may be part of the consequences of how vocabulary is handled in teaching and learning process. According to Hananto (2013), vocabulary is still neglected in the field of language teaching and learning in Indonesia. As a consequence, some studies found that Indonesian students possessed relatively low vocabulary knowledge (Kweldju, 1997; Nurweni & Read, 1999; Quinn, 1968). Thus, Kweldju (2005) argues that the Indonesian students’ single source of problem in learning English is vocabulary and suggests that the solution lies in the good handling of vocabulary.

The claim from Kweldju is not without reasons. Scholars believe that vocabulary is one of the language components that determines the success of acquiring a language. Schmitt (2000) underlines that vocabulary is central to both communicative competence and the acquisition of a second language. On the other hand, Alqahtani (2015) states that vocabulary is often considered as a critical tool for second language learners because lacking of vocabulary causes unsuccessful communication in second or foreign language. The correlation between vocabulary and language acquisition may be best described by Wilkins (1972, p. 111) on his statement: “without grammar, very little can be conveyed, without vocabulary, nothing can be conveyed.”

Nation (2006) emphasizes the importance of language learners to have sufficient vocabulary size in order to perform a second or foreign language. He estimates that in order to perform the language effectively without using any comprehension tools, we need to have 6,000 to 7,000 word-family vocabulary for spoken texts and around 8,000 to 9,000 word-family vocabulary for written texts. On the other hand, Adolphs and Schmitt (2003) estimate the need of having the most frequent 2,000 to 3,000 word families in order to participate in
basic everyday oral communication. While estimates vary, most vocabulary scholars agree that students need 3,000 to 5,000 thousand words in order to be able to read un-simplified texts with sufficient comprehension (Renandya, 2018).

In Indonesia, some studies have been conducted to figure out the number of vocabulary possessed by Indonesian students in general such as those done by Quinn (1968), Nurweni and Read (1999), and Nurhemida (2007). But, for specific students especially those majoring in English education department only few data are published. The first study was conducted by Kweldju (1997). The study measured the vocabulary size of English department students from 15 teachers’ colleges. The data showed that the participants only had 4,664 word forms or 2,800 word families. The second study was conducted by Kurniawan (2017) to assess the vocabulary size possessed by students majoring in English education department. The result of the study showed that on average the students had 1400 words. The results of the two studies revealed that the vocabulary known by the English department students was still below the threshold that is suggested by the vocabulary scholars. It may also indicate that the students are difficult to perform the English language effectively in order to cope with their daily academic tasks and communication. The results of the study may also reflect the factor that cause the students’ low level of English proficiency as previously mentioned by Ivone.

To cope with the problem, vocabulary researchers suggest the implementation of vocabulary teaching and learning strategies for both teachers and students to boost the vocabulary knowledge. In order to determine what strategies that suit the students the most and what words that students should focus on, it is necessary to diagnose the students’ present vocabulary mastery level by means of diagnostic test such as Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001). Measuring students’ vocabulary size helps teachers identify what the students lack and what words they should focus on. It is very essential to do before designing vocabulary development programs or courses in order to achieve a maximum result. Therefore, the present study aimed to investigate the English department students’ vocabulary knowledge. With this regard, the study formulated research questions as follows:

1. What is the mean of English department students’ vocabulary size?
2. What is the students’ vocabulary mastery level?
B. LITERATURE REVIEW

1. The Classification of Vocabulary

Nation (2008) classifies vocabulary into four categories based on how frequently it occurs in the language (its frequency) and how widely it occurs (its range). The four categories of vocabulary are namely high frequency words, academic words, technical words, and low frequency words. The four categories are explained as follows:

1.1 High frequency words

High frequency words are words that occur very frequently in all kinds of language uses. Those words are classified from the 2,000 most frequent word families in General Service List (West & West, 1953). They are used frequently in formal or informal situation, written and spoken text such as newspapers, conversation, novels, and academic texts (Nation, 2008). They cover 80% of the running words in most written texts and 90% of the running words in spoken texts (Nation, 2008). Most of lists of high frequency words consist of around 2,000 word families, and they are relatively short. Besides that, 169 words from the list are function words such as conjunction (and, or, but), pronoun (I, you, she, it), number (one, two, three) and so on. The rests out of the function words on the list are content words such as nouns, verbs, adjectives, and adverbs. They are also very common words in which even a very young native speaker of English seems to know them (Nation, 2008). They are required every day when we use English. These words are very important, thus Nation (2008) suggests that these words should be put on the first list of the target words in any kinds of vocabulary development program.

1.2 Academic words

Academic words are the words that occur very frequently in academic texts but they are not from the list of 1,000 or 2,000 of high frequency words. These words occur quite often in newspaper, very formal conversation, children’s books, academic writing, and other kinds of special purpose text. This group of words consist of 570 word families and are arranged in a list called Academic Word List (AWL) by Coxhead (2000). Generally the words of this list make up around 8.5% to 10% of the running words in academic texts (Nation, 2008). It means that there will be one word in every ten words coming from the AWL. Academic words do not take place very often in other kinds of language uses. It is about only less than 2% of running words or only 2 words in every 100 running words in conversation taken from the AWL. The 570 of academic words are very important for those who are using English for academic study like in universities, vocational institutes, or in schools. Thus, Nation (2008)
suggestions that the university students that use English for their academic tasks should focus on learning this vocabulary once they have mastered the high frequency words.

1.3 Technical words

Academic words are found in every kind of academic subject areas such as in Linguistics, Biology, Physics, etc. Instead of those words, there are words that are even more special to each subject area. These words are very familiar in particular areas like the vocabulary in Politics or the ones in Botany. These special words are very special to the subject area that people will recognize what subject area these words come from. These words are called technical words (Nation, 2008).

Mostly, the technical words can only be found in one specialized area. But, some words also occur in other subject areas and may have the same or different meaning. For example, the word by-pass and neck are technical words of medicine, but they are also used in some areas like Biology or Psychology. In one particular area such as in Economic textbooks which is a specialized text, technical vocabulary often appears more than usual compared to another group of specialized texts from other areas (Nation, 2008). Chung and Nation (2003) found that in anatomy about 30% of the running words are included in technical words. It means that there is approximately one word in each three is technical. Some of these technical words are available in the 2,000 high frequency words yet they are used as specialized words in a specific area. While a lot of the words seem to have distinct characteristic to a particular subject area, some would come from the Academic Word List.

For everyone who is from specialized from a specific area, technical words play a vital role. Not many of statistical studies on technical vocabulary, but probably at least 20% of the running words in high number of technical texts are possibly technical words (Nation, 2008). We have inadequate idea about the technical vocabularies but the number may vary from 1,000 words to 2,000 words according to the subject area (Nation, 2013).

1.4 Low frequency words

Low frequency words are the rest of the words in a language. They consist of a big number of word families, and even the biggest compared to the other vocabulary levels, but they rarely appear in most texts. If we count every single-word items, English may have more than 100,000 word families. Even the native speakers themselves may not know all of them. Normally native English speakers have vocabulary mastery of around 20,000 words, but it actually depends on their education level and how many readings they do in their subject areas.
or their interests (Nation, 2013). The low frequency words may occur only 2% of the running text, or, in the other words, it is only about 1 word in every 50 running words.

On the other hand, Schmitt and Schmitt (2014) assume that it is necessary to set the distinction between high frequency words and low frequency words. Regardless of academic word list and technical words, they propose a new category of vocabulary called mid frequency words as the distinction between high and low frequency words. In addition, they also revisited the number of high frequency words from 2,000 to be 3,000 most frequent word families. They then defines mid frequency words are words ranging from the most frequent 3,000 to 9,000 word families, and low frequency words are words beyond 9,000 levels.

2. The Vocabulary Levels

According to Laufer and Nation (1999), it is necessary to set the vocabulary of English in a series of levels based on the frequency of occurrence. They then suggest to list a group of 1,000 words for each level starting from the most frequent to less frequent words. One of well-known word lists is the one developed by Nation (2006). He develops a word list based on British National Corpus that makes up a fourteen word levels starting from the most frequent words to the less frequent words. The high frequency words are words from the 1st 1,000 level to the 2nd 1,000 levels (Nation, 2008) or from the 1st 1,000 level to 3rd 1,000 level (Schmitt & Schmitt, 2014). The mid frequency words are in the list between 3rd 1,000 level to 9th 1,000 level and low frequency words are in list between 9th 1,000 level to 14th 1,000 level.

The development of the word list is useful in terms of vocabulary instruction and assessment. Teachers can easily set their vocabulary focus for their vocabulary instruction by picking up from the most frequent word list. For instance, a teacher firstly needs to teach the 1st 1,000 level for his students, and once it has been mastered he should pick the 2nd 1,000 level for the next instruction focus. In fact, it is not necessary to really need to start the vocabulary development program from the 1st 1,000 word level. Instead, a teacher needs to set the vocabulary focus based on their students’ need. Therefore, it is necessary for teachers to able to measure their students’ vocabulary mastery by administering a test such as Vocabulary Levels Test (Schmitt et al., 2001). The results of the test will inform teachers about their students’ present vocabulary levels such as the levels that they have mastered and the levels that they lack of. However, teachers should focus firstly on the high frequency words and
secondly on the academic word lists if learners attempt to pursue their study at universities (Nation, 2008).

3. The Number of Vocabularies Needed to Perform the Language

One important thing that we should concern on regarding studying the size of vocabulary needed to perform the language is how large the text coverage needed to gain adequate comprehension without assisted comprehension tools such as dictionaries and translating software. In the other words, we should know how to estimate the number of unfamiliar vocabulary that can be tolerated in texts before being interfered with comprehension. The text coverage is the percentage of running words which are known by readers in a text.

Hu and Nation (2000) studied about the connection between text coverage and reading comprehension for native speakers of English by using a fiction text. They found that in a text with 80% coverage, no one got sufficient comprehension. When they use a text with 90% coverage, only a small number of the subjects got adequate comprehension. With 95% coverage or only one unknown words in every 20 running words, more subjects got adequate comprehension, and with 100% coverage of the text, most of the subjects gained enough comprehension. When a regression model was employed to the data of the study, a reasonable measure was found. The calculation found that 98% text coverage would be needed to get sufficient comprehension from reading the text. It means that there must be only one out of fifty running words that is unknown by readers. On the other hand, Laufer and Ravenhorst-Kalovski (2010) suggests two thresholds, an optimal one and a minimal one. The optimal one is the knowledge of 8,000 word families for coverage of 98%, including proper nouns. The minimal one is between 4,000 and 5,000 word families for the coverage 95%, including the proper nouns.

A study by Nation (2006) found that in the level of 98% text coverage, 8,000 to 9,000 word families are needed by readers in order to gain sufficient comprehension of written texts without using any assisting comprehension tools, and 6,000 to 7,000 word families needed for spoken texts. It suggests that readers should master all the high frequency words consisting of 2,000 word families as well as the mid frequency words consisting of 7,000 word families. On the other hand, Adolphs and Schmitt (2003) assume that 2,000 to 3,000 word families, which are high frequency words, are needed in order to participate effectively in basic everyday oral communication.
C. RESEARCH METHODOLOGY

This study employed descriptive-quantitative research design. The subjects of this study were students majoring in English education department in Kutai Kartanegara University, East Kalimantan Province. The samples of this study were selected by using convenience sampling techniques. There were 33 students from the first-year batch participated on this study. The students were employed because they were the ones available that the researcher could access. The students sit a test given by the researcher to measure their vocabulary size and mastery level. The students’ scores on the test reflected their receptive vocabulary size.

The instrument used in this study was the Vocabulary Levels Test (VLT) by Schmitt et al. (2001). The test has been widely used in vocabulary knowledge assessment (Nation, 2008) and vocabulary research (Nirattisai & Chiramanee, 2014). According to Meara (1996), VLT has been called the nearest thing to a standardized vocabulary test available. It is a diagnostic test which can also be used to measure vocabulary size (Nation, 2008). As a diagnostic test, it is used to see the students’ present vocabulary levels that can help teachers decide what vocabulary levels that learners should focus on.

The VLT that was adopted in the present study was the revised version of the one that was originally developed by Nation (1983). It consists of two blocks, one on the left and one on the right. The left block consist of six words numbered 1 to 6 as the answer choices, whereas the right block consists of 3 definitions with blanks. The test sitter should choose the available choices on the left based on the definitions on the right block by putting the number of the word chosen on the blanks.

The VLT measures five vocabulary levels based on the frequency which are the 2000, 3000, 5000, academic vocabulary (AWL), and 10000 word level. Each of those levels consists of 1,000 word families. Each level of the test consists of 30 items and it makes up the total 150 items for all levels. Each item in 2000, 3000, 5000, and 10000 levels represent 33 words (1000 : 30). It means that if a student correctly answers 20 out of 30 items, the learners know approximately 667 words from the level. For Academic Word List level, each item represents 19 words because the level only consists 570 words in total. Thus, a score of 20 out of 30 means that learners know about 380 words in the level. Nation (2008) requires 90% of the words, meaning 27 out of 30 words, should be known by learners to claim that the learners know the level. The test is available at Schmitt et al. (2001) and freely available at https://www.lextutor.ca/tests/.
To answer the research question 1, which is to investigate the vocabulary size possessed by the English department students, the results of the VLT were summarized by mean frequency and standard deviation across the five different levels (2000, 3000, 5000, 10000 and AWL). Each level represents 1,000 word families, but the academic vocabulary level which only represents 570 word families. Therefore, to get the score of the 2,000, 3,000, 5,000, and 10,000 levels, the total correct items were divided by 30 and were multiplied by 1,000. On the other hand, the score of academic vocabulary level were gained by dividing the number of correct items by 30 and multiplying it to 570. The scores of each level were summarized to get the total scores which represent the vocabulary size. The mean of the students’ score estimated the vocabulary size possessed by English department students.

To answer the research question 2, the mean scores of each level were analyzed to see whether the levels were completely mastered by the students or not. According to Nation (2008), a test sitter is considered to master a particular level of VLT if she or he correctly answers at least 90% of the total items or 27 out of 30 items in that level. The students’ mean scores on each level were analyzed to see their mastery level.

**D. FINDINGS**

Table 1 shows the results of the VLT based on each levels. The students showed the best results in 2000 level where they knew about 16 words out of 30 on average. In the 3000, 5000, 10000, and academic vocabulary, they knew about 10, 6, 2, and 7 words out of 30 respectively for each level. In total, the students knew 41 words out of 150 word families tested.

On the 2000 word level which is high frequency words, only 3 students could correctly answer 90% of the total items, there were 15 students who reached less than 50% of the items. On the 3000 word level, the highest score was 24 and there were only 2 students who could reach this level. On the contrary, 24 students, 73% of the total students, got less than 50% items correct, and even one of them could answer no items correctly. On the 5000 and 10000 word levels, the students’ scores were even worse. Most of the students on both level could only answer less than 50% correctly. On the academic words which is the group of words that is necessary for academic use, no students got 27 items correct, and the highest score on that level was only 21. Only one student got the highest score, and 7 of them got zero item correct.
Sudarman, Sumalee C. *The English Vocabulary Size And Level*

Table 1: The Results of VLT

<table>
<thead>
<tr>
<th>Levels</th>
<th>N</th>
<th>Min.</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>33</td>
<td>4</td>
<td>27</td>
<td>16.06</td>
<td>7.541</td>
<td>535</td>
</tr>
<tr>
<td>3000</td>
<td>33</td>
<td>0</td>
<td>25</td>
<td>10.24</td>
<td>6.782</td>
<td>341</td>
</tr>
<tr>
<td>5000</td>
<td>33</td>
<td>0</td>
<td>21</td>
<td>5.7</td>
<td>6.401</td>
<td>190</td>
</tr>
<tr>
<td>10000</td>
<td>33</td>
<td>0</td>
<td>9</td>
<td>2.24</td>
<td>3.000</td>
<td>75</td>
</tr>
<tr>
<td>AWL</td>
<td>33</td>
<td>0</td>
<td>21</td>
<td>6.94</td>
<td>6.230</td>
<td>132</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>-</td>
<td>-</td>
<td>41.18</td>
<td>-</td>
<td>1273</td>
</tr>
</tbody>
</table>

*Note. There are 30 questions each level and 150 questions in total. N=33*

Figure 1 shows the percentage of the students’ mean scores on each vocabulary level. Nation (2008) assumes that one is considered to master the particular levels of the VLT if she or he answers at least 90% of the items on each level. It means that a student is considered to have mastered a level on the VLT if he has a minimum of 27 correct items on the level. Therefore, the students’ vocabulary mastery level was judged based on the assumption. Figure 1 shows that none of the levels had been mastered by students because the students’ mean scores on all levels were below 90%. On the 2000 word level, the students’ mean score was 54%, or on average the students only knew about 16 words out of 30 on the test. And, only 3 students could reach the mastery level. On the 3000 word level, students knew about 34% of the items tested which was only about 10 out of 30 items. On the academic word level, the students only knew about 7 words out of 30 on average.
E. DISCUSSION

1. The English Vocabulary Size

The first research question of the study concerned on the number of word families known by the English department students in the study. The scores of the VLT reflected the participants’ vocabulary size which was assumed to highlight their vocabulary proficiency after learning English for at least 6 years, 3 years in junior high schools and 3 years in senior high schools. The results revealed that the students did better in the 2000 level, which is high frequency words, than in the other levels. Based on the findings, the English department students knew approximately 1,273 word families. This indicates a lower vocabulary size compared to the previous study by Kweldju (1997) who found that the average vocabulary size of English department students was 2,800 families. On the other hand, the results of this study are in line with the study conducted by Kurniawan (2017) who found that the first-year English department students only knew approximately 1400 words on average. The results of those studies were not different from the studies of other fields conducted by Quinn (1968), Nurweni and Read (1999), and Nurhemida (2007). Quinn investigated the vocabulary size of his first year students and found that they had known less than 1000 high frequency words, even after studying more than 6 years at schools. Nurweni and Read (1999) measured the vocabulary knowledge of the first-year students with different education backgrounds and revealed that the students had about 1226 word families. Almost the same result was also revealed by Nurhemida (2007). The result of her thesis showed that the senior high students only had about 1240 word families. By looking at these results, it can be concluded that the low vocabulary size still becomes a major problem of Indonesian students for different education backgrounds and levels.

The result of this study implied that even English department students had a low vocabulary size which indicated that they had a low level of vocabulary proficiency. Their vocabulary size did not reach the threshold as suggested by vocabulary scholars. The result may indicate that they could not perform the language effectively. As mentioned by Adolphs and Schmitt (2003), the need of 2,000 to 3,000 word families is necessary for basic daily oral communication or the need of 3,000 to 5,000 words in order to read un-simplified texts with sufficient comprehension as agreed by vocabulary scholars (Renandya, 2018). With only 1,273 word families, the students may not be able to participate in basic conversation and may find difficulties in dealing with their academic tasks. This result supports what Kweldju (2005) claims that the vocabulary is the single source of problem in learning English for Indonesian
students. Furthermore, this low vocabulary size may be the factor that contribute to the low proficiency of Indonesian students, especially the English department students as mentioned by Ivone (2005). But, indeed to prove it the researchers should study the direct correlation between the two variables in the future study.

2. The English Vocabulary Mastery Level

The data on Figure 1 shows the percentage of the students’ mean score in the VLT. The figure also indicates the vocabulary mastery level of the students. Based on Nation’s assumption, there were no levels that were completely mastered by the students. Even for the high frequency words, 2000 word level, the students were estimated to have approximately 535 out of word families or only about 54% which is very low. On the other hand, these kinds of words are the most important words that the students should have because they occur very frequently in any kinds of language use. According to Nation (2008), the high frequency words cover up to 80% in running words for written texts and 90% for spoken texts. The word level is also needed to perform an effective basic oral communication. Therefore, vocabulary scholars have suggested that ESL/EFL learners should firstly pay attention on mastering these words (e.g. Nation, 2008).

On one hand, as university students, the English department students are required deal with academic tasks. Nevertheless, the results of this study also shows that they only mastered about 23% out of 570 word families of the academic word list. It falls far below the mastery threshold for the level as assumed by Nation (2008). The result indicates that the students may find difficulties in dealing with academic tasks especially when they are required to deal with reading and listening tasks. The condition is perhaps worse since they also did not master the high frequency words. Therefore, these two issues should be put on the first priority in teaching or learning vocabulary program.

F. CONCLUSION

The results of this study show that vocabulary still becomes a major problem of English language teaching and learning in Indonesia. The English department students as one of the pillars determining the success of English pedagogy in the future had low level of vocabulary size, only 1,273 word families, in which it might share effect to their low level of English proficiency. In addition, as university students they did not even master the high frequency words as well as the academic vocabulary which are highly required for them in order to use the language and to deal with their daily academic tasks. Thus, this issue should be the
concern of English language teaching practitioners and researchers to seek for the vocabulary learning and teaching strategies that are effective to boost the students’ vocabulary size and levels. The researcher also suggests that the future research should focus on developing the students’ vocabulary size, especially on the high frequency vocabulary and academic vocabulary levels.

REFERENCES


Sudarman, Sumalee C. *The English Vocabulary Size And Level*


Nurhemida. (2007). *The relationship between morphological awareness and English vocabulary knowledge of Indonesian senior high school students.* (Master Unpublished master's thesis), The University of Queensland, St Lucia.


