THE WRITTEN TEXTS USED IN FIRST-SEMESTER COURSES IN THE ENGLISH LANGUAGE EDUCATION STUDY PROGRAM

Laurentia Sumarni

Dosen Program Studi Pendidikan Bahasa Inggris, FKIP, Universitas Sanata Dharma
Alamat korespondensi: Kampus I Mrican, Jl. Affandi, Yogyakarta.
Email: *laurentia.sumarni@gmail.com*

ABSTRAK

Riset telah menemukan bahwa untuk dapat memahami teks secara memadai, diperlukan pemahaman atas setidaknya 95% dari kosa kata yang ada dalam teks. Di Pendidikan Bahasa Inggris (PBI) Universitas Sanata Dharma, pemerolehan kosakata difasilitasi melalui pembelajaran kosakata incidental. Namun beban kosakata dari masing-masing teks belum diukur untuk memastikan bahwa teks tersebut mengandung kata-kata yang memiliki frekuensi yang tepat, penggunaan yang luas, dan penyebaran yang merata untuk memastikan text coverage. Oleh karena itu, riset untuk mengetahui text coverage dan beban kosakata dalam teks yang digunakan dalam teks perkuliahan di semester satu Pendidikan Bahasa Inggris akan dilakukan.

Instrumen riset yang digunakan adalah tes pengenalan kosakata yang diujikan kepada 64 mahasiswa semester satu. Tes tersebut terdiri dari empat (4) teks yang berisi 100 kata yang diambil dari teks yang digunakan dalam mata kuliah Basic Reading dan Book Report. Evaluasi beban kosakata yang terdapat dalam teks yang digunakan dalam mata kuliah semester satu dilakukan dengan menggunakan the Corpus of Contemporary American English online. Enam teks dianalisa dengan menggunakan corpus.

Hasil survey menunjukkan bahwa secara rata-rata mahasiswa tidak mengenali sekitar 3,62 kata yang terdapat dalam teks. Maka, text coverage mahasiswa adalah 96,38%. Ini berarti bahwa mahasiswa siap membaca teks-teks bacaan yang diberikan pada semester satu karena kosa kata yang ada di dalam teks-teks tersebut masih berada pada jangkauan kosakata mereka. Secara rata-rata, kata-kata yang tidak dikenali mahasiswa adalah 6,75 kata yang termasuk dalam daftar RANGE 2 dan 44,25 kata yang termasuk dalam daftar RANGE 3. Analisis teks COCA menunjukkan bahwa teks yang dianalisa mengandung 65,8% kata yang termasuk RANGE 1, 13,8% kata yang termasuk RANGE 2, 20,3% kata yang termasuk RANGE 3, dan 6,3% kata yang termasuk kosakata ACADEMIC. Dapat disimpulkan bahwa mahasiswa semester satu sudah siap dengan materi yang diberikan pada perkuliahan semester satu. Oleh karena itu, sangat direkomendasikan apabila semua teks bahasa Inggris yang akan dipakai sebagai materi perkuliahan dianalisa terlebih dahulu dengan menggunakan COCA untuk memastikan bahwa kosa kata yang diberikan telah sesuai dengan level mahasiswa. Dengan demikian, kesuksesan pembelajaran bahasa Inggris dapat dipastikan.

Kata kunci: *text coverage, vocabulary load, vocabulary size, vocabulary knowledge, incidental vocabulary learning, COCA*

1. INTRODUCTION

Vocabulary is the currency in language learning. Knowledge of the vocabulary in the text, or the lack of it, will affect comprehension. To be able to understand an English text, it is necessary to know at least 95% of the words in the text. It means that in every 100-word text, there are only 5 unfamiliar words. When students' text coverage is under 95%, chances are that students will find the reading text difficult to understand. Nation (2006) states that "if 95% coverage is required, a vocabulary size of the most frequent 4,000 word families may be necessary for comprehension of novels and newspapers. If 98% coverage is required, knowing 8,000 word families may be necessary to understand newspapers, and

knowing 9,000 word families may be required to understand a novel." High text coverage will guarantee effective and successful learning and will in turn increase the vocabulary size.

In the English Language Education Study Program of Sanata Dharma University, vocabulary acquisition is facilitated through incidental vocabulary learning in such courses as Book Report, Basic Reading, Basic Listening, Intermediate Reading, Intermediate Listening, and so on. Explicit vocabulary learning is facilitated in the Vocabulary Course in the second semester. Since the students' vocabulary size ranges from 3,500 to 9,500 words, the students' coverage of the text (the number known words) varies from one student to the other. The failure to understand the text will result in poor grades and students' vocabulary acquisition is not facilitated. Reading for pleasure will turn into reading for torture because students do not have enough vocabulary to understand a certain text. Students cannot enjoy a text if there are more than 10 unknown words per 100-word text. Unless addressed appropriately, this problem which is initially a vocabulary problem will change into a motivation problem because students' lack of vocabulary results in learning failure.

So far, a research on text coverage has never been done to investigate the appropriateness of texts and to investigate how many unknown words students encounter in the text. This kind of research will be essential for planning learning materials to suit the students' vocabulary learning needs. However, vocabulary has not been used to determine the suitability of texts to their levels. There are no existing documents stating the required vocabulary size for each level. Therefore, it is necessary to investigate students' percentage of coverage of the existing texts used in the courses in the ELESP, Sanata Dharma University. This research will help the curriculum designers to prioritize the high-frequency words over low-frequency words.

In addition, teachers have difficulty in selecting appropriate texts to be used in the class. It is a hard task for them to estimate whether students have already reached 95% of text coverage and it will be time consuming to conduct a text-coverage research every time the text is going to be used. Therefore, the use of the online Corpus of Contemporary American English (COCA) is expected to help teachers to evaluate the vocabulary loads of the texts that they use in the classrooms. The evaluation of vocabulary loads will contribute to the appropriate selection of text materials and in turn will facilitate better and more effective language learning program. Therefore, this research aims to investigate the percentage of text coverage among the first-semester students of the English Language Education Study Program and to evaluate whether the written texts used in the first-semester courses have contained vocabulary load which is graded according to their frequency of use, range of use and dispersion.

2. VOCABULARY IN TEACHING AND LEARNING

For the purpose of this research, it is important to define vocabulary and underline its significance in English language teaching and learning. Lehr, Osborn, & Hiebert (2004) cited in Read (2000) define vocabulary as "words we use to communicate in spoken and written language. Receptive vocabulary refers to the words we understand through reading and listening, while productive vocabulary refers to words we use to communicate through writing and speaking" (as quoted in Hanson and Padua, 2011).

The crucial role of vocabulary in learning and communication is undeniable. In the courses in a university level, students are exposed to various types of texts containing vocabulary, ranging from academic, general high-frequency level, to specialized low-frequency level vocabulary. Students' success in understanding texts depends highly on their vocabulary size and knowledge. The bigger the size of their vocabulary, the better they will understand the texts, thus the greater success they will experience. Laufer & Sim (1985) cited in Webb & Nation, found that "vocabulary knowledge may be the best gauge of whether or not a text will be understood. Research indicates that for learners to be able to guess words in contexts and gain adequate comprehension of written text it is necessary to know at least 95% of the words (Laufer 1989).

In the English Language Education Study Program, vocabulary acquisition is facilitated through incidental vocabulary learning in such subjects as Book Report, Basic Reading, Basic Listening, Intermediate Reading, Intermediate Listening, etc. Through these subjects, students are expected to acquire vocabulary through the intensive study of words in the courses. In Book Report, for example, students are required

to read English novels. In Basic Reading, students are exposed to simple texts; while in Basic Listening, students are exposed to an audio recording containing short speeches and descriptions. The texts are chosen based on the goals of the course that students must achieve by the end of the course. The choice of materials is not based on vocabulary load of the texts and the vocabulary burden of the course, but on the topics and course goals.

Before proceeding further, the type of vocabulary the learners need to know must be clarified first. According to Nation (2008:7), "vocabulary can be divided into three or four levels largely on the basis of how often it occurs in the language (its frequency) and how widely it occurs (its range)." The types of vocabulary include:

1) High frequency words
   This is the most important group of common words which occur very frequently and are used in formal and informal occasions, in speech and writing, and in novels, conversation and newspapers. This group consists of 2,000 word families, which are mostly short, and 169 of them are function words. The rest of the words are content words (noun, adjective, adverb, and verb). According to Nation (2008:8), "in most texts around 80% or more of the running words are from the most frequent 2,000 words of English."
2. Academic words
   Coxhead (2000) lists 570 word families, called the Academic Word List. The characteristics of academic vocabulary are:
   a) A group of words which is not from the most frequent 1,000 or 2,000 words;
   b) is frequent and widely used within a certain specialized area;
   c) typically make up around 8.5% - 10% of the running words in academic texts, approximately 1 word in every ten comes from the Academic Word List.
3. Technical words
   These are the words with even more special purposes and these are the words that are very common in one particular area, such as the vocabulary of Economics or the vocabulary of Agricultural Engineering. Almost one in every three words are technical words (Nation, 2008:10), which are also in the 2,000 high frequency words but have a specialized use in a particular area. Statistically, at least 20% of the running words in most technical texts are likely to be technical words (Nation, 2008:10). Technical vocabularies range in size from around 1,000 words to 5,000 words depending on the subject area.
4. Low frequency words
   It is a diverse group of words, consisting of (a) words that are not quite frequent or wide range enough to be high frequency words, (b) technical words from other areas, (c) words that just occur rarely. If only single-word items are counted, English probably has a low frequency vocabulary of over 100,000 word families. According to Nation (2008:11), "in friendly conversation, about 5% of the running words are low frequency words, in newspapers they are about 10%, and in academic texts they are about 10%."

Teachers must know these types of vocabulary so that they develop materials in the light of this knowledge. Ideally, all materials are adjusted to the students' current vocabulary size and the targeted vocabulary size intended by the course. Good texts are those who facilitate the vocabulary acquisition. Nation (2001) states that "comprehension and incidental vocabulary learning through reading are likely to increase if the percentage of known words in a text is 98%." It means that students will likely gain vocabulary meaningfully if they can recognize 98 words per 100-word text. He further asserts that "if 95% coverage is required, a vocabulary size of the most frequent 4,000 word families may be necessary for comprehension of novels and newspapers. If 98% coverage is required, knowing 8,000 word families may be necessary to understand newspapers, knowing 9,000 word families may be required to understand a novel."

In light of these facts, teachers must consider students' vocabulary size and the texts' vocabulary load in order to know the text coverage. To know students' vocabulary size, a vocabulary size test, either customized or established tests can be used to measure students' current vocabulary size. However, to decide which texts should be included as teaching materials is a challenging task. Inappropriate texts may lead to students' lack of understanding of the texts and learning failure. Therefore, it is urgent for

teachers to evaluate the vocabulary load of the written texts they use in the classrooms to ensure maximum text coverage. Some online tools can be utilized to evaluate the vocabulary load of written texts, namely the British National Corpus (BNC) and the RANGE program. Another tool is called the Corpus of Contemporary American English (COCA). In this research, the vocabulary load will be evaluated using COCA.

The Corpus of Contemporary American English (COCA) is the largest freely-available corpus of English, the only large and balanced corpus of American English. It was created by Mark Davies of Brigham Young University, and it is used by tens of thousands of users every month (linguists, teachers, translators, and other researchers). The corpus contains more than 450 million words of text and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. The interface allows the user to search for exact words or phrases, lemmas, part of speech, or any combinations of these. The corpus also allows the user to easily limit searches by frequency and compare the frequency of words, phrases, and grammatical constructions. COCA allows the user to (a) determine the vocabulary size necessary to understand the vocabulary in text, (b) create word lists based on the frequency of occurrence and range of use of vocabulary in different types of discourse, (c) determine the number of encounters with words in a text, and (d) to evaluate the vocabulary load of text for teaching and learning language. With this facility, teachers can easily analyze words in the texts that they will use in the classrooms to see whether they facilitate vocabulary acquisition and comprehension.

There is a close relationship between vocabulary size and text coverage. It is hypothesized that when students' vocabulary size is big, the text coverage (the number of known words in a given text) will be bigger. Thus, comprehension is facilitated.

According to Nation (2008), in order to deal with a range of unsimplified spoken and written texts require learners to have at least 8,000 word families. Nation (1990, 1993, 2001) further adds that "the critical importance of developing an adequate high-frequency vocabulary since learner's skill in using the language is heavily dependent on the number of words they know, particularly in the early stages of learning a foreign language, with around 3,000 word families being a crucial threshold." Hirsch and Nation (1992) argue that "in order to reach text comprehension, readers need to be familiar with 95% of the words in a text." This text coverage percentage is logical because "to cope well in English, a second language learner would need around 5,000 words and preferably 10,000 words" (Nation, 2004).

The acquisition of high-frequency word families is urgent to ensure text coverage. Francis and Kucera (1982) suggest that "the 2,000 most frequent word families of English make up 79.7% of the individual words in any English text, the 3,000 most frequent word families represent 84%, the 4,000 most frequent word families make up about 86.7%, and the 5,000 most frequent word families cover 88.6%. A much better reading comprehension power is ensured if a reader knows the meanings of at least 90% of the words in a text." In line with that, Schmitt states that "the vocabulary in the 2,000-3,000 frequency band provides additional material for spoken discourse, but additionally, knowledge of around 3,000 word families is the threshold that should allow learners to begin to read authentic texts. Most research indicates that knowledge of the most frequent 5,000 word families should provide enough vocabulary to enable learners to read authentic texts." In short, the researchers agree that the bigger vocabulary size, the bigger the text coverage and the better learners will understand spoken and written discourses. Therefore, teachers should strive to provide learning materials in the form of written texts which are appropriate to the current vocabulary size and targeted vocabulary size so that maximum text coverage is achieved and successful learning is attained.

Considering the importance of vocabulary knowledge in order to understand English texts, a survey on text coverage and an evaluation of vocabulary load in the written texts are urgent to be conducted. Hirsch and Nation (1992) argue that 'knowledge of the vocabulary in the text is one of the many factors that affect reading.' In addition, vocabulary load in the written texts needs to be considered to guarantee the suitability of the written texts. Webb (2007) states that 'the number of encounters with unknown words may provide some indication of their potential learning.' However, too many unknown words in the texts will not facilitate learning. Hirsch (1992) asserts that 'having to struggle with reading because many words are unknown will take a lot of pleasure out of reading."

Thus, it must be decided how much of the vocabulary in the text needs to be known. As previously understood, too many or too few unknown words will not facilitate learning. Many researchers suggest that knowing the students' current vocabulary size will enable teachers to determine which texts to be used to help them acquire more words. Previously, the relationship between vocabulary size, text coverage and word list has been discussed by Waring and Nation(1997). Hirsh hypothesizes that 'if a reader knows 90% of the running words (tokens) in a text, then there will be 10 unknown words in every ten. If each line in the text contains about 10 words, then there will be one unknown word in every line.' Comprehension will be difficult at this stage. Waring and Nation (1997) argue that in the light of students' vocabulary size, text coverage of 90% requires students to know at least 6,000 word families.

Once the students' vocabulary size has been known, teachers can select materials that will be suitable for the students. Teachers can use the COCA to evaluate the vocabulary load of the written texts they will use in the classrooms. With this theoretical foundation, the survey on text coverage and the evaluation of vocabulary load of the written texts used in the first-semester courses will be conducted.

3. RESEARCH METHODOLOGY

This research involved a survey on text coverage and evaluation of vocabulary load in the written texts. The participants of the survey were the 64 first-semester students, while the subjects of the evaluation were the written texts used in the first-semester courses in the English Language Education

Table 1: The Number of Unfamiliar Tokens per 100 Tokens and the Number of Lines of Text Containing One Unfamiliar Word (Hirsch and Nation, 1992)

% Text Coverage	Number of Unfamiliar Tokens per 100 Tokens	Number of Text Lines per 1 Unfamiliar Word
99	1	10
98	2	5
97	3	3.3
96	4	2.5
95	5	2
94	6	1.6
93	7	1.4
92	8	1.25
91	9	1.1
90	10	1

The table shows the relationship between the percentage of text coverage and the number of unfamiliar tokens and the number of text lines per 1 unfamiliar word. The text coverage of 99% implies that there is only 1 unknown word per 100-word text and the unknown word can be found in every ten lines in the text. This text will be easy for learners to understand. However, the text coverage of 90% implies that there are 10 unknown words in every 100-word text and the unknown words can be found in every line in the text. This text will be very difficult for the learners to understand and reading becomes a burden. This research tries to investigate what is the learning burden learners have to take in every text. The information of students' average vocabulary size is important.

Study Program. The texts were randomly selected from the Basic Reading, Book Report, and Pronunciation Practice Courses. To collect the data of the text coverage survey, the researcher created sampling of the written texts used in the first semester. The results were a series of 100-word texts divided into 6 types. Each type contained 4 sets of 100-word texts. These texts were tested to 64 first-semester students to find out their text coverage by asking them to circle the unknown words in every 100-word text. The results were counted quantitatively and tabulated to see the average text coverage of the written texts.

Then, the evaluation of vocabulary load of the written texts used in the first-semester courses was conducted using the Corpus of Contemporary

American English. First, the written texts were inputted into the COCA online. The results were examined to see the vocabulary load of each written text used. After the survey and evaluation, the report was written by drawing conclusions and giving recommendations.

4. RESEARCH FINDINGS AND DISCUSSION

This section will present the results of the research both qualitatively and quantitatively. This section will provide the research results to answer the two questions. To answer the first question, i.e. *what is the percentage of text coverage among the first-semester students of the English Language Education Study Program?*, a survey was conducted. The survey was meant to test students' text coverage by distributing 6 different types of tests randomly. Each student did one type of text coverage questionnaire. Each type contained 4 sets of 100-word texts taken randomly from texts used in the first-semester courses. The texts were taken from the Basic Reading, Book Report and Pronunciation Practice Courses. The texts from Basic Reading were "Old Ways, New World", "Diversity and Tolerance." The texts from Book Report Course were Christmas Carol by Charles Dickens, Sing to the Dawn by Minfong Ho, and The Adventures of Tom Sawyer by Mark Twain. Finally, the text taken from Pronunciation Practice I was "Speech Mechanism." However, the texts "Diversity and Tolerance" and "Speech Mechanism" were not tested. From the texts, a 100-word set of texts was arranged. Then, the sets were distributed in 6 types of texts to be tested. The following table shows the distribution of texts. The complete description of the tests and texts can be seen in Appendix 1 (page 26) and Appendix 2 (page 38).

Then, the students were asked to circle the unknown words in the text. Afterwards, the texts were collected for analysis.The following table shows the complete results. The table will be discussed in detail. Table 4 is the summary of the students' unknown words and the average text coverage in the six types of tests.

Table 2: The Distribution of Texts

Type	Text 1	Text 2	Text 3	Text 4
1	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*
2	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*
3	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*
4	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*
5	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*
6	100 words from *Old Ways, New World*	100 words from *Christmas Carol*	100 words from *Sing to the Dawn*	100 words from *Adventures of Tom Sawyer*

Table 3: The Average Students' Text Coverage

The Number of Unfamiliar Words in Texts					
Type	Text 1	Text 2	Text 3	Text 4	
1	3.38	6.46	2.38	3.92	4.038
2	2.09	11.36	1.27	1.81	4.13
3	3.45	5	3.72	3.09	3.81
4	2.9	4.4	4.3	6	4.4
5	1.88	4	3.22	0.55	2.41
6	2.5	3	4.3	2.1	2.97
	2.7	5.8	3.1	2.9	3.62

As it can be seen in Table 4, for students doing the Type 1 Set, they found in average 3.38 unknown words in Text 1, 6.46 unknown words in Text 2, 2.38 unknown words in Text 3, and 3.92 unknown words in Text 4. The average unknown words were 4.038. Thus, students' text coverage was 95.96%. It means that in average, students found 4 unknown words in each text. However, each text did not have the same level of difficulty. Text 2 was considered the most difficult text as it students found in average 6.46 unfamiliar words.

For the Type 2 Set, students' average text coverage was 95.87% because they found 4.13 unfamiliar words in average. The most difficult text was Text 2 in which students found 11.36 unfamiliar words in average. However, Texts 1, 3, and 4 seemed to be easy for them as they did not find too many unfamiliar words. For the Type 3 Set, Type 4 Set, Type 5 Set and Type 6 Set, the average was 96.19%, 95.6%, 97.59%, and 97.03%, respectively. The most difficult set was Type 4, Text 4 where students found in average 6 unfamiliar words in the text. Types 5 and 6 were the easiest types because students only found in average 2 unfamiliar words in every text.

In terms of the texts' level of difficulty, Text 2, i.e. Christmas Carol, was considered the most difficult text. It was understood because the text was the unabridged version of the novel, while the Adventures of Tom Sawyer was the simplified version of the original novel, and Sing to the Dawn was a novel intended for teenagers in Asia. Therefore, the vocabulary was adjusted to the students' level.

From the data and discussion, it can be concluded that the students' average text coverage was 96.38%. It means that students found in average 3.62 unfamiliar words in every 100-word text. Therefore, in average, students did not find difficulties with the first-semester texts.

The unfamiliar words that the students found in the texts are listed in Table 4.

Table 4: The List of Unfamiliar Words in the Texts

TEXT NUMBER	RANGE 2	RANGE 3
	501 – 3000	> 3000
Text 1	ancient, challenges, engineer, gained, marriages, particularly, pursue, route, software, strike (10)	Caste, arranged, tension, abolished, anger, core, defy, delicate, divorced, elsewhere, forcefully, freewheeling, gradually, headscarves, humiliating, impression, ironclad, irreconcilable, limbs, pervasive, prevailing, priests, prospective, rare, resignation, scholars, spurning, stature, submissive, tensions, wed (31)
Text 2	beneath, freshest, glanced, hung, sheet, spring (6)	chuckle, ragged, scarcely, shaving, amends, amongst, anxious, appalled, awful, bedpost, berries, blaze, bore, cab, ceiling, chimney, chinked, conversed, crisp, degradation, dread, dumb, encompass, enthroned, exceeded, exclaimed, forth, glared, gleaming, glistened, grove, hearth, holly, horrible, humility, ivy, lurked, meagre, menacing, merchants, mistletoe, obedience, perversion, petrification, pinched, prostrate, recoiled, recompensed, roaring, scattered, scowling, seals, shreds, shrivelled, shuddering, stale, sticking-plaster, thoughtfully, till, tints, trifled, undergone, wolfish (63)
Text 3	bent, firm, glanced, leaning, rung, stared, storm (7)	ankle, cage, darted, dawn, sparrow, abbot, ached, afforded, bare-headed, braided, caged, clinging, clumsily, cocked, dais, daze, dragged, dreaded, drenched, empty-handed, glimpse, glimpses, gloomy, hobbling, jerked, ladder, monks, monsoon, mud, murmured, pattering, pine, rearranging, ripping, rutted, searing, solitude, speckling, splintered, steadily, streaming, sunk, thrust, timidly, unblinking, veranda, weave, wipe (48)
Text 4	proved, slip, slipped, sprang (4)	cave, inn, alley, bunch, lantern, wharf, widow, abyss, astonishment, bats, blindfold, burglar, cavern, ceased, crept, darting, depths, descended, distant, faint, flocking, fright, glinting, gloom, innermost, knots, mad, murmur, mutter, pitch-black, recesses, thunder, tin, vast, weary (35)

Table 4 shows the unfamiliar words students found in the texts. Most of them are included in RANGE 3, which means that the words are listed in the COCA list from 3000 words and above. It can be concluded that students were already familiar with high-frequency words and found difficulty in understanding less frequent words. Some students were still unfamiliar with words belonging to RANGE 2 (501-3000).

To answer the second question, i.e. *Have the written texts used in the first-semester courses contained vocabulary load which is graded according to their frequency of occurrence, range of use and dispersion?,* the Corpus of Contemporary American English was used to analyze the texts.

The texts being analyzed using COCA were 4 texts previously tested to the students to find their average text coverage, namely "Old Ways, New World", "Christmas Carol", "Sing to the Dawn" and "The Adventures of Tom Sawyer." Two additional texts were included. They were "Diversity and Tolerance" and "Speech Mechanisms." The text "Diversity and Tolerance" will be labeled as Text 5, and "Speech Mechanisms" will be labeled as Text 6. The texts were inputted in COCA and the results can be seen as follows.

Table 5: The Frequency Range of the Texts

TEXT NUMBER	WORD COUNT	FREQUENCY RANGE			
		RANGE 1	RANGE 2	RANGE 3	ACADEMIC
		1 - 500	501 - 3000	> 3000	
Text 1	701 words	68%	13%	19%	9%
Text 2	604 words	70%	10%	20%	3%
Text 3	639 words	65%	14%	21%	0%
Text 4	657 words	68%	13%	19%	0%
Text 5	728 words	63%	16%	21%	14%
Text 6	834 words	61%	17%	22%	12%
	65.8%	13.8%	20.3%	6.3%	

Table 6 shows that the texts ranged from 604 to 834 words. All of the texts contained more than 60% of words in RANGE 1 (between 1 – 500). The words in RANGE 2 were around 10 – 17% of the total words, and 19 – 21 % of the total words contained words in RANGE 3 (> 3000). Meanwhile, the first two texts contained ACADEMIC words between 3 to 9% of the total words, while the second two texts did not contain ACADEMIC words. The last two texts contained ACADEMIC words between 12 and 14 % of the total words. In average, 65.8% of the words belonged to RANGE 1, 13.8% of the words belonged to RANGE 2, 20.3% of the words were included in RANGE 3, and the ACADEMIC words made up the 6.3% of the total words. The complete result of the analysis can be seen in Appendix 3 on page 50.

Table 5 also shows that the words belonging to RANGE 3 were bigger in size than the words belonging to RANGE 2. It implies that students found new and unfamiliar words in the texts. Words in RANGE 2 should be bigger in size than the words in RANGE 3.

5. CONCLUSIONS AND RECOMMENDATION

This section will discuss the conclusions of the research results on text coverage and vocabulary load and offer some suggestions and recommendations to follow up this research. The discussions will be presented according to the order of the research questions.

First,to find out the percentage of text coverage among the first-semester students of the English Language Education Study Program, a survey was conducted to 64 first-semester students. The survey was done by distributing 6 types of vocabulary recognition test. Each type contained four 100-word texts. The texts were taken from the Basic Reading and Book Report courses. The students were asked to circle the unfamiliar words they found in the texts. The results showed that in average, students found 3.62 difficult words. Therefore, the students' text coverage was in average 96.38%. The most difficult text was Text 2 (Christmas Carol) which was an unabridged novel, in which students in average found 5.8 difficult words.

It can be concluded that students found more unfamiliar words in unabridged novels and found less unfamiliar words in simplified novels and reading texts. This means that the students are ready to face the reading texts in the first semester as the vocabulary was within their range.

Second, to find out whether the written texts used in the first-semester courses have contained vocabulary load graded according to the frequency of occurrence, range and dispersion, the online Corpus of Contemporary American English was used to analyze the texts. The texts being inputted in the software were 6 texts, four previous texts and 2 additional ones. The results showed that in average, 65.8% of the words belonged to RANGE 1 (0-500), 13.8% of the words belonged to RANGE 2 (501 – 3000), 20.3% of the words were included in RANGE 3 (>3000), and the ACADEMIC words made up the 6.3% of the total words. The result of the analysis confirmed the fact that the students were already familiar with high-frequency words (the first 3000 words). The unfamiliar words the students identified belonged to words in RANGE 3 (>3000). In average, students found 6.75 unfamiliar words belonging to RANGE 2, and 44.25 unfamiliar words belonging to RANGE 3 out of 100 words. Considering the gap between the text coverage (96.38) or 3.62 unfamiliar words in a 100-word text and 44.25 unfamiliar words identified by all students, it means that the words which are unfamiliar to one student may be familiar to others, and vice versa.

Based on the results, it can be concluded that in average the students of the English Language Education Study Program were already familiar with 96.38% of the words being used in the first-semester courses. The materials were also appropriate and challenging to be used in the first-semester courses.

To ensure students' success in understanding texts in the courses, it is recommended that the materials be analyzed first using COCA to see whether they are appropriate to be used according to the students' vocabulary size. When the vocabulary is within students' range, it is guaranteed that the learning will be beneficial for students.

REFERENCES

Anderson, R.C., & P. F. 1983. Reading comprehension and the assessment and acquisition of word knowledge. In B. Houston (ed.), *Advances in Reading/Language Research*. Volume 2, pp. 231-256. Greenwich, CT:JAI Press.

Aziez, F. 2011. "Examining the Vocabulary Levels of Indonesia's English National Examination Texts". *Asian EFL Journal,* Professional Teaching Articles, Vol. 51, April 2011.

Cervatiuc, A. 2008. "ESL Vocabulary acquisition: Target and Approach". *In the Internet TESL Journal,* Vol. XIV, No. 1, January 2008. Available from *http://iteslj.org/Articles/Cervatiuc-VocabularyAcquisition.html*.

Coombe, C. *Assessing vocabulary in the language classroom*. In Malaysian English Language Teaching Association, pp. 111- 124.

Coxhead, A., 2010. "Grabbed early by vocabulary: Nation's ongoing contributions to vocabulary and reading in a foreign language". In *Reading in a Foreign Language*, April 2010, Vol. 22, No. 1, pp. 1-14.

Francis, N., & H. Kucera. 1982. Frequency analysis of English Usage. Boston: Houghton Mifflin Company.

Hanson, S., & Padua, Jennifer F. M. 2011. *Teaching Vocabulary Explicitly*. Honolulu: Pacific Resources for Education and Learning (PREL).

Hirsch, D., & Nation, P. 1992. "What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure?" In *Reading in aForeign Language*, Vol. 8 (2), 1992.

http://corpus.byu.edu/coca/

Huang, J., Yuedong, Sheng., & Ying, X. 2007. "On the Vocabulary Size for Chinese English learners". In *Polyglossia,* Vol. 13, October 2007.

Meara, P. 1996. The Vocabulary Knowledge Framework. In *Lognostics.*

Meara, P., & Alcoy, J. C. O. 2010. "Words as Species: an Alternative Approach to Estimating Productive vocabulary size". In *Reading in a foreign Language*, April 2010, Vol. 22, No. 1, pp. 222-236.

Nagy, W.E., & Scott, J. A. 2000. "Vocabulary Processes". In M. L. Kamil, P.B. Mosenthal, P.D. Pearson. R. Barr (Eds.) *Handbook of Reading Research: Volume III,* pp. 269-284. Mahwah, NJ: Erlbaum.

Nation, I.S.P. & Beglar, D. 2007. "A vocabulary size tes"t. *The Language Teacher, 31*(7), 9-13. Nation, I. S. P. (1990). Teaching and Learning vocabulary. New York: Newbury House.

Nation, I.S.P. 1995-1996. "Best Practice in Vocabulary Teaching and Learning". *EA Journal,* 3 (2), 7-15.

Nation, P., & Waring, R. 1997. "Vocabulary Dize, Text Coverage and Word Lists". In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press, pp. 6-19.

Nation, I.S.P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I., & Heatley, A. 2002. Range: A program for the Analysis of Vocabulary in Texts [Computer Software]. Available from *http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx*

Natin, I.S.P., & Beglar, D. 2007. "A Vocabulary Size Test". In *The Language Teacher*, 31(7), pp. 9-13.

Nation, I.S.P. 2008. *Teaching Vocabulary: Strategies and Techniques.* Boston: Heinle Cengage Learning.

Nurweni, A., & Read, J. 1999. "The English Vocabulary Knowledge of Indonesian University Students". *English for Specific Purposes*, 18, pp. 161-175.

Pikulski, John J., & Templeton, S. 2004. *Teaching and Developing Vocabulary: Key to Long-Term Reading Success*. Houghton Mifflin Company. Retrieved January 9, 2013, from *http://www.eduplace.com/state/pdf/author/pik_temp.pdf.*

Read, John. 2007. "Second Language Vocabulary Assessment: Current Practices and New Directions". *International Journal of English Studies*. Vol. 7 (2), 2007, pp. 105-125. *www.um.es/ijes.*

Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N., & Meara, P. 1997. "Researching Vocabulary Through a Word Knowledge Framework: Word associations and Verbal Suffixes". In *SSLA,* Vol. 20, pp. 17-36. Cambridge University Press.

Schmitt N., & McCarthy, M. 1997. *Vocabulary, Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

Schmitt, N. 2000. *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Schmitt, N. 2008. *Teaching Vocabulary.* Pearson Education, Inc.

Scott, Judith. A., Hoover, Merrit., Flinspach, Susan Leigh., & Vevea, Jack L. (n.d.) ––––. "A multiple-Level Vocabulary Assessment Tool: Measuring Word Knowledge Based on Grade-Level Materials". In *National Reading Conference Yearbook,* No. 57, pp. 325-340.

Thorndike, E., & Lorge, I. 1944. *The Teachers' Word Book of 30,000 Words*. Columbia: Teachers College.

Watts, J. ––––. Measuring the Effectiveness of a Vocabulary Programme. Pp. 15-25.

Xing, P., & Fulcher, G. 2006. *Reliability Assessment for Two Versions of Vocabulary Levels Tests*. *www.sciencedirect.com.*