

## **BAB I**

### **INTRODUCTION**

This chapter covers the background of the research, statement of the research problem, objective of the research, significance of the research, scope and limitation of the research and definition of key term.

#### **A. Background of the Research**

Corpus in plural form, refers to collection of text, written or spoken, which are organized systematically according to the purpose of its construction. Corpus investigation have been already used in many general linguistics and also especially in English linguistics even in many different aspects of language teaching researches. Linquist (2009:03) also stated that W. Nelson Francis and Henry Kucera at Brown University were the first pioneer, they compiled 500 samples of 2.000 words each from a selection of various text genres in present-day American English, and the collection of texts was commonly known as Brown Corpus.

The existence of Brown Corpus inspired the others linguists to conduct the same study. The Brown Corpus later followed by arose a number of other corpora, such as the Lancaster-Olso/Bergen (LOB) Corpus, which used the same format as the Brown Corpus and made it possible to compare different varieties of English (Grigaliuniene, 2013:09). According to Leech (1992) cited by Kholkovskaia (2017) these two corpora became the model for other comparable corpora of the so called “Brown Family” that contained of seven corpora covering the period between 1931-2006.

There were numerous corpus based study that conducted in various particular language workings, such as fiction, magazines, conversations, newspapers, academic papers, literatures and so on.

Moinnvaziri (2012) tried to investigate vocabulary used in general English textbook and used a computer program called Vocabprofile (VP). The reason for the researcher to select the textbook because it was the only textbook which is taught for General English course nationally which has been taught for more than ten years. The result of this study show that the vocabulary presented in the reading passages of the textbook is mostly suitable for beginner and low intermediate levels but the level of textbook cannot be an efficient one for the university students. And for the frequency of words presented in the textbook does not seem to provide enough instances of new items and neither does it provide enough opportunity for practicing these items. The only vocabulary items which are acceptably rehearsed are the words of first 1000 frequent words which are not suitable and challenging for the level of the students.

Ling (2016), she conducted a corpus analysis in magazine horoscopes. She tried to investigate the use of modal auxiliary verbs of the horoscopes. For this study, the corpus consisted of horoscope from 108 issues of 3 magazines namely CLEO, Female, and the Malaysian Women's weekly from the year 2011-2013. From the result of this study, Ling found that the most frequent word which appear in horoscope is pronoun "You" with 5.01 % of total occurrences. Then, the most frequent modal auxiliary verb used in horoscope was modal of probability "will" and followed by "May". Modal "will" was counted 76.46 % on its occurrences of total modal verbs occurrences in the corpus. It could be concluded that the writers of horoscope try to predicts and state the possible future events on their horoscopes.

Kartal (2017), a study conduct to determine the most frequent adjectives used in academic texts and to investigate whether these adjectives differ in frequency and function in social sciences, technology, and medical sciences. This study focuses on the academic texts of The Corpus of

Contemporary American English (COCA), a contemporary and genre-based corpus. COCA was used for the research because it is free to access, very comprehensive and highly representative data which includes over 450 million words.

Moon (2016) from University of Washington, She studied the similarities and differences in word use between The Time of London and Kenya's Daily Nation. The researcher used the news sections of The Times and the Daily Nation during 2013 which all the content indexed by Factiva. The study focused on the equivalent comparison across outlets and excluded the most irrelevant items such as recipes, music reviews, opinion columns, and wire stories. Moon compared the two word lists by its parts of speech, specifically for the pronouns, adjectives, nouns, and verbs in its top 500 words, then examined the characteristics of the two corpora which represent the news style of press systems in different political and social environment. From the analyses, Moon agreed that the two news publications, the Daily Nation and The Times, show some key differences that indicate stylistic and topical variation. Greater word frequency of both politics and geography were generally devoted in the Daily Nation, and it tends to investigate the topics in terms of internal news and institutional power, whereas more space of international and global news and citizens were devoted in The Times.

Mazdayasna and Firouzi (2013) conducted a study to investigate the frequency distribution and syntactic functions of adjectives across literary and technical texts. The data were collected from two corpora, namely novels and engineering academic books in English. The result of the study shows that there is a significant difference in the frequency use of adjectives across the two corpora. From a register perspective, the high frequently use of adjectives in technical texts in comparison to literary texts (67.3% and 32.7%, respectively).

Kusumaningrum, Maria Asumpta Deny. (2014). Conduct to a Analysis of Vocabulary Use of a Junior High School Textbook. this study was designed to look throughout at the vocabulary coverage and word recycling in a Junior High School textbook of *English in Focus* grade seven. The textbook was published by the government. There are two problems to be solved in this research: (1) what is the vocabulary coverage of Junior High School textbook? (2) How the words are recycled. In this research, the researcher applied corpus based analysis. There searcher collected the data from the textbook obtained the electronic book from <http://bse.kemdikbud.go.id>. The researcher used RANGE and the FREQUENCY program in order to analyze the data. The program, created by Nation, is available at [http://www.vuw.ac.nz/lals/staff/Paul\\_Nation](http://www.vuw.ac.nz/lals/staff/Paul_Nation). The RANGE program was used to answer first research problem, while the FREQUENCY program was used to answer the second research problem. The findings showed that all of the chapters have not reached the appropriate level vocabulary coverage of 95%. Chapter I has the highest level of vocabulary coverage which covers 94.85%. It means that chapter I is the easiest of all chapters for students because 94.85% of tokens found in most 2,000 frequent English words. On the other hand, glossary has the smallest result of vocabulary coverage which covers 46.81%. The result of word recycling showed that only chapter VIII has the highest occurrence numbers in more than ten times. Overall result of the whole textbook, it needs to be supplemented with other materials containing more word families and sufficient number of recycling of each word family, based on students' needs, characteristics, background knowledge, and wants. However, the textbook provides a lot of activities which are familiar to the students.

Sun (2016) focused on the analysis of vocabulary coverage in children story series "The Magic Three House". The researcher used a corpus analysis

approach to analyze the vocabulary coverage within a single-author story series, its repetition of vocabulary, and the incidental and intentional vocabulary learning opportunities therein. Then, the result of word list from The Magic Three House compared with three word list published by The Ministry of Education in Taiwan. This three word list is a guideline for book developers to follow. In particular, the findings indicate that the more basic the vocabulary list is, the higher the percentage of repetition. Therefore, extensive reading of only one story series will not only expose learners to a large portion of words from the lists, but many of them repeatedly

Herdian, Damasus Desta. (2017). Focus in A Corpus Analysis of Vocabulary Coverage and Word Frequency of Junior High School Course Books. This research was aimed to discover the characteristics of the vocabulary coverage and word frequency in Junior High School course books, grade Eight. With two research questions, which are: (1) what is the vocabulary coverage of Junior High School course books in terms of types, tokens, and word families? (2)How frequent the words are being used in the course books? The research method was corpus-based survey. The data were obtained from four junior high school course books, published by official/government and private publishers. Two Paul Nation's concordance programs and a modified program (based on Bauman and Culligan's wordlists) were used as instruments for data analysis. They were RANGE and FREQUENCY. The findings showed that the four course books have average 30,463 tokens, and 3,063 types. The course books covered 72.55% tokens from total first 800 words (GSL800) vocabulary which should be mastered as Junior High School learners. The course books covered average 622 out of 800 word families or 77.78% of the total 800 headwords (GSL800) as required for the 2nd grade of Junior High School learners. It is considered not enough and need some improvements. The contents of the course books are

accessible enough to students whose vocabulary knowledge is within the range. Unfortunately, they have inadequate number of word families and students will find difficulties in producing speaking and writing (active English) unless they use several supplementary English books or worksheets. The course books do not meet the criteria of incidental learning and text coverage either the learning opportunity to the students is not really high. The average deficit vocabulary is 177 out of 800 headwords. Almost 54% word families from the course books are repeated less than ten times. Thus, it gives small opportunities to students for deepening learning vocabulary. There are 20 functional words and 4 content words in the total 24 most frequent word types from all course books. The words *the*, *to*, *a*, and *and* are four most frequent words appeared in all course books. The course books have most of the types, tokens, and families listed in the GSL wordlist. However, authors need significant improvement in designing the material to give students a higher opportunity in learning vocabulary. High amount of tokens, types, and word families are also necessary but should be followed with suitable vocabulary needs and high frequency of essential words to help students become more effective and efficient in learning vocabulary.

From some of the studies above, the study of corpus is growing rapidly along with the development of computer technology. Various linguistic studies and the field of language using the corpus, and build in different language and size based on the purpose of the study. McCarthy (2004:02) corpus appears for many languages and complies for different purposes. The corpus is used in dictionary setting, reference material, vocabulary, grammar and study material and also book language course.

Considering some previous research in the field of vocabulary coverage, this research also discusses vocabulary coverage. In this research, the researcher will discuss vocabulary coverage in Junior High School

textbook entitled English in Focus. From the explanation above, it is clearly seen that vocabulary takes an important role in studying English especially for students. It is because they cannot avoid English although they have graduated from Junior High School or even from university. Facing English all the time, students are familiar with vocabulary. It means that they should always improve their size of vocabulary in order to follow their English need. If they could not improve their vocabulary size, they would not be able to follow their English lesson well.

Considering the problem above, the researcher is going to look at the vocabulary of English Junior High School textbook. This research is based on a corpus study. This study does not attempt to judge of textbook quality, but to find the vocabulary used in the textbook, with the hope that future studies will improve on the methodology used. This research can also be used as a way to help in analyzing the overall quality of textbook and help to create better English material for students.

According to the research above, only a few researcher have tried to examine nouns that are often in textbooks. Therefore, at this present study the researcher wanted to research about the most frequently used noun in course books. Research about of the most frequent verbs can be used to improve the students' vocabulary mastery. Wilkins (1972) stated that vocabulary knowledge is crucial to successful communication because vocabulary plays an important role in delivering meaning. It means, vocabulary is needed to make us able to understand the meaning of we are spoken or listen about In a well communication, the students are mastering nouns, adverbs, and adjectives vocabulary only is not enough to make their communication in English well. They need to knowledge how noun are used in communication

This research is analyzing one course books of junior high school to find the most frequently used noun. The researcher is interested in conducting a research entitles “**A Corpus Analysis of The Most Frequently Used Noun in the English in Focus of Junior High School**”

#### **B. Statement of the Research Problems**

Based on the background of the study above, the questions that formulated by the researcher is:

1. What are the most frequently used noun in the course book of junior high school?
2. What are the word frequency of the course books?

#### **C. Objectives Of The Research**

Based on the research question above, the purpose of the study is to know the most frequently used noun in the course book of junior high school

#### **D. Significant of The Research**

The researcher expects that this study will give contribution for:

1. Student

Students will get understand the nouns most often used in course books to help students in understanding groups of people, animals, or things

2. English Teachers

Teachers can provide an authentic material based on the most frequently used noun in the course book. So that, teachers are able to give an example about the proper noun to describe an object



### 3. The other researchers

The result of the study can be used as the reference for another researcher to conduct a research with similar study or to find the deeper, more detailed result and also be able to investigate and correct the shortages from the previous study.

## E. Scope and Limitation of the Research

To avoid deviations from the purpose of the study, this study only focuses on analyzing the nouns most often used in textbooks. Textbooks downloaded from website. Researcher just choose textbooks aimed at junior high school <https://annibuku.com/bse/buku-english-in-focus-kelas-8-smp-70> and <http://bse.kemdikbud.go.id>. The weakness in this study is that it only uses 1 textbook and 3260 words. So for further researchers hopefully they can provide a large number of textbooks and build a bigger corporation.

## F. Definitions of the Key Terms.

1. A corpus is a collection of texts, written or spoken, that represents language in the real use, and usually stored in a computer database (McCarthy, 2004: 1).
2. The course book is Matsuoka and Hirsh (2010) defined a course book (or course book) as a book which is designed specifically to be used in a classroom language learning environment with language teacher support including pre-teaching of target vocabulary items. Therefore, course book then refers to any kind of book which is used by teachers to support English language learning in a classroom. The data were obtained from four junior high school course books, published by official/government and private publishers

3. Frequency

According to Nation (1983), most frequency count is based on a sample of text with at least one million words

4. Nouns are commonly defined as words that refer to a person, place, thing, or idea (Evelyn P. Altenberg and Robert M. Vago 2010 )