

CHAPTER III

RESEARCH METHOD

This chapter presents research design, data and data sources, technique of data collection, and data analysis.

A. Research Design

The purpose of this study is to find out the most frequent nouns that are often used in course books. This research uses corpus as only the data analyzed, the study was conducted using corpus based study. According to Kaewphanngam et.al. (2002), is how to analyze a collection of languages that occur naturally in form both recorded speech and written text. Corpus is referred to samples of language that stored electronically in a computer. With modern corpus software, researchers can do identify the use of the most common and least common items in a large corpus the amount of data and helps researchers to choose the most frequent and most numerous unusual creative items for research.

this study, researchers used the corpus as the only data and corpus data in written form. The corpus is composed of one course books of English in focus of junior high school, the course book was published by government. In analyzing the corpus, researchers use some basic tools in the corpus such as: frequency list or word list, concordance, etc. According to Grigaliuniene (2013: 43) a frequency list is a list of all the words that appear in a corpus and can be ordered alphabetically or by frequency to identify the most and the least general items used in corpus. Concordance is a type of search engine to look through the corpus and list every occurrence of words or

phrases. In modern times, there is a lot of software or free access freeware to analyze the collection in which each software provides tools such as list of frequencies and concordances.

The research began by selecting textbooks then, errors and foreign languages are edited or deleted and followed by build a corpus. Finally, the corpus is marked and then analyzed using computer software named TagAnt 1.2.0 and AntConc 3.5.2 developed by Laurence Anthony from Waseda University. Corpus data in this study were obtained from the website.

B. Data And Data Source

1. Data

Data is all information used to answer research questions. Because this research uses a corpus based study, the corpus is the only data used. In this study, researchers chose one textbook to find the most frequent nouns from the course books and raise research objectives. All data to be analyzed is collected from textbooks that report facts or information about the world. The textbook selected by researchers in this study is only one textbook entitled English In Focus.

2. Data source

Data source is source that used by the researcher for collecting the data. The researcher chose one course books entitle English in Focus to use in the corpus data in this study. The course books was taken from website (<https://annibuku.com/bse/buku-english-in-focus-kelas-8-smp-70> and <http://bse.kemdikbud.go.id>)

This website has a lot of course books which certainly authentic and could be analyzed to find the most frequent words of the course books

C. Technique of Data Collection

In this study, the researcher used one technique of collecting the data, which was documentation. Documentation can be in the form of texts, pictures, records, or somebody's works. Bogdan and Biklen (2007:133) say that documentation refers to the collecting data or materials such as autobiographies, newsletters, diaries, newspapers, books, articles, and so on. In this study, the researcher compiled the document in the form of texts, specifically the document are in the form of course books. The course books was taken from a website. To find the authentic data for the study, the researcher was carefully choosing the appropriate website. After that, course books are taken freely from the internet. They are stored in the form of Microsoft Word documents to create researchers check for errors or useless language in the text easily. Then, selected course books are arranged into corpus.

D. Data Analysis

Data analysis is an essential process where the researcher faces massive amounts of field notes, interview transcripts, audio recordings, video data, reflections, or information from documents, all of which must be examined and interpreted. According to Shamoo and Resnik (2003) various analytic procedures "provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data. In this study the process of analyzing data included: the first researchers download and choose course books from the Internet. Choose a website that provides a complete collection course books. Some web sites provide files in PDF format (Portable Document Format), but some of them are not in PDF format. If the

resulting file is not in the form of a PDF file, it must be copied and saved

Course books in the form of word document files. After that, check one one by one to make sure all words are copied.

Before analyzing the corpus using software, corpus files are converted to TXT file format to make data readable, because the corpus analysis software can only read the txt file format. That Corpus files are converted using Notepad ++ v7.5.6, a free software program developed by the Free Software Foundation and can be accessed at <https://notepad-plus-plus.org>. After the text is converted, the entire content from corpus in txt file format is checked whether there are some errors or foreign languages that need to be edited. After a mistake, an error is unchecked, corpus has been tagged using TagAnt to mark parts of all speeches words and make noun analysis can be done quickly and accurate. Corpus analysis using AntConc 3.5.2. AntConc 3.5.2 is free software, a multiplatform tool specifically created for corpus analysis developed by Laurence Anthony of the Center for English Language Education at Japanese Waseda University Science and Engineering. By using this software, researchers can generate a list of frequencies from all words corpus course book. The corpus that has been tagged part of his speech analyzed using the Cluster / N-gram tool at AntConc to produce a list of the frequency of words that have been marked as objects. From the process of using Cluster / N-gram tool, the results will appear in which form of table contains a list of words used in the corpus and their total rank and frequency, and they can be sorted from the most frequently used to the least often.

The result of the Anconc and TagAnt software are then divide into several categories.