CHAPTER III

METHOD OF THE STUDY

3.1. Research Approach

In this study, the writer uses the mixed methods design which consists of qualitative and quantitative approach to analyze the data. According to Dörnyei (2007:163), a mixed methods study involves the collection or analysis of both quantitative and qualitative data in a single study with some attempts to integrate the two approaches at one or more stages of the research process. The writer uses mixed methods design in this study because the research process integrates quantitative and qualitative approaches. The quantitative method is used in the calculation of the frequencies of the collocations. This is done automatically by the computer program in the corpus website. The result is a numeric data together with the collocations. However, the collocations which are calculated by the computer program still needs to be selected by using the qualitative method. This is because the computer program sometimes presents inflections as different items, while they should be considered the same. Consequently, the adjectival collocates of the word 'beauty' can only be determined correctly by combining the quantitative and qualitative methods in this study.

3.2. Population and Sample

The population in research does not always mean of a number of people. Population can also consist of objects, people or even events (e.g. miners, schools, revolutions). Based on that definition the population of the writer's study is the words from COHA and COCA. Both corpora were created by Mark Davies of Bringham Young University. COCA contains more than

410 million words of text and is the largest freely-available corpus of American English. COHA contains 400 million words and is balanced in each decade for the genres of fiction, popular magazines, newspaper, and academic prose (Lee, 2010: 113).

From the description about both corpora, the sources are too big to be analyzed. That is why the writer needs to take a sample. The sample which the writers take is based on the purposive sampling method. According to Babbie (2007: 184), purposive sampling is a type of nonprobability sampling in which the units to be observed are selected on the basis of the researcher's judgment about which ones will be the most useful or representative. The writer takes a sample which consists of ten periods of time which is each period has a span of 10 years. The judgment or the consideration is related to the objective of this study which is to compare the changes of noun collocates of 'beautiful' from 1911 to 2010. The further explanations are presented in the following section.

3.3. Technique of Data Collection

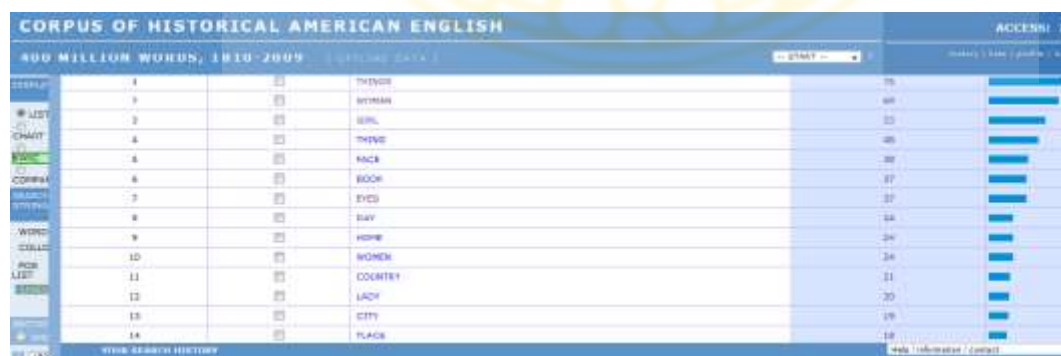In collecting the data, the writer used four steps. First, the writer opened the website www.americancorpus.org. In this study, the data have been collected and recorded in COCA and COHA by Mark Davies of Bringham University. In order to use corpora well, the writer needed to log in with registered email and password. Second, the writer needed to sign in the website. Then, the writer was able to see the webpage of COCA as shown in Figure 3.1

Figure 3.1 The screenshot of COCA

Third, the writer went to the left panel as shown in figure 3.1, then the writer went to the

search string and put the word 'beautiful' in the WORD(S) box. Then the writer chose the part of

speech, that is noun.ALL which refers to all noun, in the POS LIST box. After that the writer

chose number 1 for left box and 2 for right box of the number option in the COLLOCATES,

which means that the writer searched for 1 words before the adjective 'beautiful' and 2 words

after it. The writer chose the time by blocking the years from 1911 until 1920 in the left box and

click IGNORE in the right box. Then, the writer chose frequency in the sorting part to see how

the results would be sorted, where the most frequent result appears first. Finally, the writer

clicked the search button. And the result is shown in the figure 3.2.

Figure 3.2 The screenshot of the result

Fourth, the writer needed to select the appropriate nouns collocate. As shown in the

figure 3.2, the result of the noun collocates in the word 'beautiful' in the years 1911-1920 are

THINGS, WOMAN, GIRL, THING, FACE, BOOK, EYES, DAY, HOME, WOMEN, COUNTRY, LADY, CITY, PLACE. Then, the writer needs to select the top 10 nouns because there are some of the nouns shown which are not suitable for the writer's study. For the example, the writer exclude the word THINGS and WOMEN because they occur two times. So, the top 10 appropriate nouns collocates are WOMAN, GIRL, THING, FACE, BOOK, EYES, DAY, HOME, COUNTRY, LADY. Then, the writer did the same things for the next period of time.

3.4. Technique of Data Analysis

There are three steps which are the writer used in the technique of data analysis section. First, the writer needed to make a table listing the topten collocates in each of the 10 periods of time: 1911-1920, 1921-1930, 1931-1940, 1941-1950, 1951-1960, 1961-1970, 1971-1980, 1981-1990, 1991-2000, 2001-2010. These 10 periods of time include a time span of a hundred years which comform to the standard of a dichronic study.

Second, the writer grouped the nouns into their hypernymy. This means that the words which are considered hyponyms are combined into a more general category. According to Matthews (1997: 167), hyponymy is the relation between two lexical units in which the meaning of the first is included in that of the second. In order to make the work more consistent, the writer used WordNet to determine the hyponym and the hypernym. The WordNet is a large lexical database of English which is developed by Princeton University. WordNet is containing the definition of words, but it is organized conceptually because of the basic unit in WordNet is a synonym set (synset) which represents a lexicalized concept.

Third, the writer interpreted the result. This was done by relating the results with several references, including U.S History, People Magazines, etc. In interpreting the results, the writer

used the corpus driven approach, instead of the corpus based approach. According to McEnery and Hardy (2012:6) a corpus based approach typically use corpus data in order to explore a theory or hypothesis, typically one established in the current literature. A corpus driven approach rejects the characterization of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language. Therefore, the interpretation may be considred assumptions based on the evidence from the corpus.