

CHAPTER III

METHOD

3.1 Research Approach

In this study, the writer uses mixed-methods approach to complete the analysis of language and gender in fairy tales. According to Fischler (2011), mixed methods approach is a procedure for collecting, analyzing, and mixing both quantitative and qualitative research and methods in a single study to understand a research problem. Creswell (2003, p. 20) also emphasizes that the data collection also covers gathering both numeric information as well as text information in order that the final database represents both quantitative and qualitative information. This approach is appropriate for this study since the writer is going to use numerical data which is obtained through computer software, *Antconc*; and to explain the data collection in order to see the concept of male and female representation through the adjective and verb collocates in the *Hans Christian Andersen Fairy Tales*.

The approach will be applied in analysing the language used in *Hans Christian Andersen Fairy Tales*. As the nature of corpus linguistics study, the quantitative part of this study is the frequency of word or phrase which can be calculated automatically using *Antconc* as computer software. Nevertheless, the result cannot be analyzed loosely; it should be filtered further. There are some qualitative processes in ascertaining and finding certain collocates as it will explain in chapter IV later. As a result, the adjective and verb collocates that

attach to male and female characters in *Hans Christian Andersen Fairy Tales* can only be examined by combining the quantitative and qualitative model in this study.

3.2 Corpus

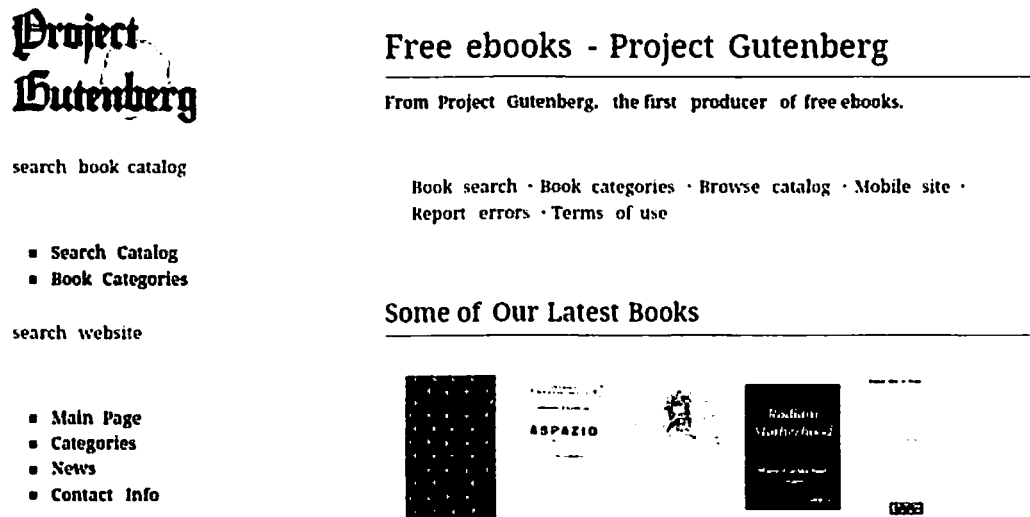
The corpus of this study consists of documents, particularly *Hans Christian Andersen Fairy Tales*. This collection is chosen because this famous fairy tales book has been known for ages and has been translated into more than 150 languages included Indonesian. There are 128 fairy tales which contains approximately 373.400 words. It is published under copyright of The Floating Press in 2010 and edited by The Team of Floating Press. The version of *Hans Christian Andersen Fairy Tales* is in PDF format. Therefore, it needs to be converted in Plain Text Format so that it enables the writer to process it using computer software *Antconc*.

3.3 Technique of Data Collection

There are three steps taken in collecting the data namely; downloading files; converting files; and processing files in *Antconc*. As mentioned previously, the corpus consists of *Hans Christian Andersen Fairy Tales*. This fairy tales can be downloaded for free by accessing website of Project Guttenberg on the address www.gutenberg.org. According to the website, Project Gutenberg is the first and the largest single collection of free electronic books or eBook that was found in 1971 by Michael S. Hart. Most of the collections are the full text of public domain

books. The collections are available in formats such as PDF, HTML, MOBI, EPUB, and Plucker.

Figure 3.1 Project Guttenberg's Homepage Screenshot



The downloaded files are in the PDF format. This file needs to be converted into .txt format so that it can be processed in the *Antconc* software. *Antconc* is a computer software which mostly used for analysing data in corpus linguistic major. The software can be installed for free without any registration. After converting the files, the next step that the writer does is using *Antconc* in order to sort the main data that only cover certain nouns namely; 'boy', 'girl', 'prince' and 'princess'. Below are the procedures of collecting the data using *Antconc*:

1. Open *Antconc* computer software by double clicking on it, then open file directory. Choose the data of *Hans Christian Andersen Fairy Tales* in .txt format.

concentrates on the adjective and verb collocates, the other word classes other than adjective and verb are deleted. The writer select appropriate adjective and verb collocates based on some considerations explained later in Chapter IV. These words are categorized based on their synonyms. The categorization is made by the help of *WordNet*. *WordNet* is a large lexical database of English, which all word classes are grouped into sets of cognitive synonyms (synsets), each expressing distinct concept.

After doing those two steps, the third step is comparing the similarities and differences of adjective collocate between 'boy' and 'girl'; verb collocate between 'boy' and 'girl'; adjective collocate between 'prince' and 'princess'; and verb collocate between 'prince' and 'princess'. Finally, the writer interprets the result which discuss further in Chapter IV.

CHAPTER IV

DISCUSSION