

CHAPTER III

RESEARCH METHOD

In this chapter, researcher presented research design, data and data source, technique of data collection, technique of data verification, and data analysis.

A. Research Design

This study was set to determine the most frequently used verb in the children short stories for junior high school students. In order to accomplish this objective of study, the researcher designed this study as a corpus based study. McEnery stated that corpus based study generally use corpus data in order to explore an already established theory or hypothesis in the current literature that it would be possible to validate, refute, or refine it (McEnery & Hardie, 2011, p. 6). In this research, the corpus data were obtained from children short stories texts as the corpus. In linguistics, corpus that is appropriate for analysis generally should consists of complete, naturally occurring text in oral or written form which the origins and source are well documented (Sinclair, 1996 in Hasko, 2012, p. 4670). In this case, the researcher chose written corpus which are selected children short stories text suitable for junior high school.

A corpus based study may belong to either qualitative or quantitative method since it often combines both approach in processing the data. It is also a characteristics of corpus linguistics to begin with quantitative result and work towards qualitative one (Leech, Hundt, Mair, & Smith, 2009).

Corpus-based study in general includes the use of a broad, representative electronic archive of spoken or written texts, or both and the use of computer-assisted analytics techniques for accuracy and efficiency. The large extent of text will be hard to be analyzed manually. In this research, the researcher attempted to calculate the frequency pattern of verbs in the selected children short stories for junior high school by utilizing text analyzer programs, namely Antconc 3.2.4 software. From this calculation, it generated quantitative data of the most occurring words which resulted into frequency list. This lists displayed the number of time each words appear in the text. In order to determine which words is the verb, the researcher also combined the analysis with a part of speech (POS) tagging tool, TagAnt. After gaining the frequency data, the result can be interpreted qualitatively in order to elaborate why a particular verb frequency pattern occurs. Therefore, it might help to develop a better understanding of the language in context, for example in selecting and identifying a suitable materials for language teaching.

B. Data and Data Source

Data is the facts or information obtained to reach the research goals. Because the aim of this research is to discover most frequently verb appears in the selected children short stories for junior high school, this study belongs to corpus based study. Therefore, the data for this study were

obtained from the corpus which is the compilation of children short stories for junior high school.

In order to obtain the data for conducting analysis, the researcher needed a lot of children short stories to be selected and compiled as the corpus data. In this study, several children short stories are selected. So, the corpus of this study consists of about 61 short stories suitable for the junior school students. In choosing the children short stories, the researcher selected a lot of already famous classics children short stories from various genre and authors that are commonly chosen by teachers for junior high school. Among those short stories are Lamb to the Slaughter by Rod Dahl, Click Clack the Rattlebag by Neil Gaiman, Seventh grade by Gary Soto, etc. Hence, the data in this study is in the form of written corpora.

Table 3.1 The Study of Written Corpora

No.	Author	Short Story Title
1	A.A. Milne	A LOST MASTERPIECE
2	Abraham Lincoln	THE GETTYSBURG ADDRESS
3	Ambrose Bierce	A HORSEMAN IN THE SKY
4	Ambrose Bierce	ONE SUMMER NIGHT
5	Amrita Pritam	WILD FLOWER
6	Amy Tan	RULES OF THE GAME
7	Anne Hart	THE FRIDAY EVERYTHING CHANGED
8	Anton Chekhov	A DEFENSELESS CREATURE
9	Anton Chekhov	A COUNTRY COTTAGE
10	Banjo Paterson	THE DOG
11	Benjamin Franklin	THE WHISTLE
12	Bret Harte	THE LUCK OF ROARING CAMP
13	Chinua Achebe	CIVIL PEACE
14	Edgar Allan Poe	THE TELL-TALE HEART
15	Edgar Allan Poe	THE CASK OF AMONTILLADO
16	Frank Stockton	THE LADY, OR THE TIGER
17	Gabriel Garcia Marquez	A VERY OLD MAN WITH ENORMOUS WINGS

18	Gary Soto	SEVENTH GRADE
19	George Ade	THE FABLE OF THE PREACHER WHO FLEW HIS KITE, BUT NOT BECAUSE HE WISHED TO DO SO
20	Giovanni Boccaccio	FEDERIGO'S FALCON
21	Guy de Maupassant	THE NECKLACE
22	Guy de Maupassant	THE HAND
23	H.G. Wells	THE TREASURE IN THE FOREST
24	H.G. Wells	THE STAR
25	H.H. Munro (SAKI)	THE OPEN WINDOW
26	H.H. Munro (SAKI)	CLOVIS ON PARENTAL RESPONSIBILITIES
27	H.H. Munro (SAKI)	THE LUMBER ROOM
28	H.H. Munro (SAKI)	THE DREAMER
29	Jack London	TO BUILD A FIRE
30	Jamaica Kincaid	GIRL
31	Junius Edwards	LIARS DON'T QUALIFY
32	Katherine Mansfield	THE FLY
33	Kurt Vonnegut, Jr.	HARRISON BERGERON
34	Langston Hughes	ONE FRIDAY MORNING
35	Laura E. Richards	MAINE TO THE RESCUE
36	Laura E. Richards	THE GOLDEN WINDOWS
37	Liam O'Flaherty	THE SNIPER
38	Lucille Fletcher	THE HITCHHIKER
39	Mark Twain	A MONUMENT TO ADAM
40	Marta Salinas	THE SCHOLARSHIP JACKET
41	Matsuo Basho	THE AGED MOTHER
42	Maya Angelou	WHEN I LAY MY BURDEN DOWN From I KNOW WHY THE CAGED BIRD SINGS
43	Nancy Bell	A CLEVER THIEF
44	Nathaniel Hawthorne	THE HAUNTED MIND
45	Neil Gaiman	CLICK CLACK THE RATTLEBAG
46	O. Henry	ARISTOCRACY VERSUS HASH
47	O. Henry	HEARTS AND HANDS
48	O. Henry	LORD OAKHURST'S CURSE
49	O. Henry	THE GIFT OF THE MAGI
50	O. Henry	THE RANSOM OF RED CHIEF
51	Oscar Wilde	THE CANTERVILLE GHOST
52	Oscar Wilde	THE SELFISH GIANT
53	Philip K. Dick	THE EYES HAVE IT
54	Richard Connel	THE MOST DANGEROUS GAME
55	Roald Dahl	LAMB TO THE SLAUGHTER
56	Sherman Alexie	VALEDICTION
57	Shirley Jackson	CHARLES
58	Stephen Leacock	MY FINANCIAL CAREER

59	Stephen Leacock	THE NEW FOOD
60	Virginia Driving Hawk Sneve	THE MEDICINE BAG
61	W. W. Jacobs	THE MONKEY'S PAW

In collecting the data, a source for retrieving the corpus is needed. The researcher chose online source for the data. In modern research, online source is more useful and effective because it can be accessed anytime and anywhere. There are a lot of source for retrieving short stories online. In this study, the researcher only selected the ones that are appropriate for certain age and level of study especially junior high school students. The 61 stories for analysis were compiled and downloaded from online corpora database. In this study, the researcher determined <https://americanliterature.com/middle-school-short-stories> as the main data source. The secondary data source for this study are online websites providing alternative information that could not be obtained from the main data source. Those websites offer additional information and reference for this study such as:

- <https://www.weareteachers.com/best-short-stories-for-middle-schoolers/>

The researcher also used journal, article, and books to provide more reliable information and could be added to complement the reference of this study.

C. Technique of Data Collection

In gathering the research data, one or more types of technique are required. In order to achieve the purpose of the analysis, the researcher used one data collection tool, which was the documentation. Documentation can

be in the form of documents, images, records, or works by other scholars. As discussed by Bogdan and Biklen (2007:133), the documentation refers to the compilation of data or materials such as autobiographies, newsletters, diaries, journals, books, posts, etc. In this analysis, the researcher collected the document in the form of texts, in particular the document in the form of short stories. The stories have been taken from the website. To determine the authentic data for the analysis, the researcher carefully selected the appropriate website. After that, the texts that could be used for the study can be stored or downloaded in the form of word documents text. The texts contain 61 titles of the children short stories which have been freely taken from the Internet. They are stored in the form of a Microsoft Word document to make it easy for the researcher to verify the error or useless language in the text. The selected papers were then compiled to be a corpus but it required to be reviewed one by one first to make sure that all the words are copied before they are computerized.

D. Data Verification Technique

Regarding this analysis trustworthiness, the researcher needed to make sure that the data is valid and authentic. In order to validate the analysis first, the researcher need to carefully check spelling or grammatical errors repeatedly. So that the analysis in the electronic analytic software could not be obstructed. After the errors was checked and corrected, the researcher utilized TagAnt, a simple freeware tagging tool to find the most commonly

occur verbs hence the validity of verbs tagged is checked. TagAnt can verify the verbs because it tagged the part of speech (POS) of each word of the text before the corpus evaluated using AntConc. Therefore, the researcher could confirm every words contained verb to be collected as corpus in the next procedure of analysis. TagAnt will tag any words as its part of speech or words classes such as whether it's noun, pronoun, an adverb, an adjective, a verb, etc. After tagging using the tagAnt, the researcher sorted and evaluated further words in the short stories corpus entirely that have been tagged as a verb.

E. Data Analysis

Data analysis is a critical method in which researchers are faced with vast quantities of data collection and examination and make use of it to analyze and interpreted data. Anderson and Arsenault (1998:138) stated data analysis is a systematic method that organizes data into usable units, integrates and synthesizes concepts, establishes constructs, themes, patterns or theories and sheds light on the important results of the study. In this research, the data analysis method included: compiling the data from online source, reviewing the stories one by one to make sure all the words are copied in order to compiling the data into a corpus, converting the data into TXT files to make the corpus readable in a software tool, Checking spelling and grammar error in each words so the analysis in the tools could be done easily, tagging the part of speech of each words hence the most frequent

verbs appears could be calculated, analyzing the data using selected tools: Antconc 3.2.4, and last drawing on the interpretation.

The first procedures of analysis was compiling the data to be analyzed therefore it could be constructed as corpus. The researcher downloaded the short stories from the online source. In selecting the websites source, the researcher considered suitable websites that include a full selection of short stories or provide short stories based on the age and English level of students. Some websites offered the file only in readable format like PDF that is not editable or corrected when there were errors in the text, but some of them were also in word document format which can be edited. If the file in the form of a PDF file, the researcher would copy and save it in the form of a microsoft word document file or .doc so it could be converted into plain text format that more suitable for the toolkit or text analyzer. The researcher used notepad program in order to convert the word file into TXT or plain text format. The researcher also checked the all content of corpus text carefully that the words not contained error for simplifying the analysis. Any error or mistakes in the language that were not necessary were reduced or corrected.

After the corpus was constructed in the form of plain text (TXT format), the researcher used a part of speech tagging tools to decide which words were verbs. The researcher used TagAnt software that facilitate in tagging each words of corpus into its own classes of part of speech such as

noun, adjective, adverb, verb, etc. to obtain the corpus that contained the verbs therefore the further analysis could be done easily.

After the corpus that contained tags of verbs were gained, the researcher then calculated the percentage of verbs that appear in the corpus by using the Cluster / N-grams system in AntConc to generate a frequency list of terms that were tagged as verbs. From the procedure of using the Cluster / N-grams tool, the findings were a table containing the words list of corpus and the calculation of its occurrence in the text, that can be ordered from the most frequent to the least frequent one.

As a result of the AntConc word frequency list, the researcher was able to see the most often used verbs in children's short stories for junior high school. The researcher then picked the top 100 verb list to be analyzed. The list of verbs was copied in Microsoft Office Excel and sorted by the frequency, after which all the percentages of the verbs were counted. After the list of verbs frequency has been generated, the researcher categorized each verb according to Beth Levin's (1993) English verb classes theory. The verbs divided into at least forty-three verb classes according to the semantic meaning (Levin, 1993)

From there, the researcher could determine which verb category that most frequent than the others or dominates the texts. Lastly, after the result of the analysis were found to answer the research question, the researcher then creates an interpretation so that it can be used in drawing the conclusion of the study.