CHAPTER II

REVIEW OF RELATED LITERATURE

In this chapter, the researcher presents the result of reviewing related literature which covers Corpus based analysis, children short stories, verbs, and the previous studies.

A. Corpus Based Analysis in Children Short Stories

In these last five decades the work that takes the concept of using corpus has been increased. Corpus, the plural forms are certainly called as corpora, refers to the collection of text, written or spoken, which is systematically gathered. A corpus can also be defined as a broad, principled set of naturally occurring examples of electronically stored languages (Bennet, 2010, p. 2). For such studies, corpus typically refers to a set of authentic machine-readable text that has been selected to describe or represent a condition or variety of a language (Grigaliuniene, 2013, p. 9). Likewise, Lindquist (2009) also believed that corpus is related to electronic. He claimed that corpus is a collection of texts stored on some kind of digital medium to be used by linguists for research purposes or by lexicographers in the production of dictionaries (Lindquist, 2009, p. 3).

Nowadays, the word 'corpus' is almost often associated with the term 'electronic corpus,' which is a collection of texts stored on some kind of digital medium to be used by linguists for research purposes or by lexicographers for dictionaries. McCarthy (2004) also described corpus as a

collection of written or spoken texts, typically stored in a computer database. We may infer from the above argument that computer has a crucial function in corpus (McCarthy, 2004, p. 1).

In this regard, computers and software programs have allowed researchers to fairly quickly and cheaply capture, store and handle vast quantities of data. The benefits of using computers in corpus linguistics are substantial (Leech in Mcenery & Gabrielatos, 2006, p. 34). Therefore, the linguist can analyze more material and get more precise frequency measurements by using a corpus (Lindquist, 2009, p. 5).

Aside from that, the text used in corpus is systematically composed. It means the corpus is chosen on the basis of the purposes in building the corpus in some structure and material. Hunston (2006) said that the text is typically chosen to represent a kind of communication or a variety of languages (Hunston, 2006). For instance, in the early 1960s, Nelson Francis and Henry Kucera formed the first electronic-based corpus at Brown University, and it is intended to represent the use of language in various text genres in present American English.

Linguists have collected a wide range of corpora for various purposes, and the number of corpora is increasingly rising throughout the years. Each corpora has a different text genre, and different word number size. Corpora could be different in size, collected for various reasons and assigned to different text types (Hunston, 2006). In some degree, corpora may be completely homogeneous, or heterogeneous. Corpora which consists of all

homogeneous texts from one language or a single register, etc. Whereas a corpora that is all heterogeneous is constructed from a variety of texts. In addition, for particular purposes, corpora can be built in various sizes. A very broad corpora, for example, may be used to construct a dictionary.

According to Lindquist (2009: 10) eight forms of English language corpora exist:

a. Spoken Corpora

Spoken language is one extent that is frequently understated in general corpus. It is because spoken corpora are technically complex and costlier to compile. In fact, majority of people talk more than they write and less read than listen to others more speaking. Spoken corpora generates transcripts of the spoken speech, such as transcripts of ordinary conversations captured in individuals' homes and workplaces or telephone calls, corporate meetings, TV programs or radio broadcasts (McCarthy, 2004, p. 1). Spoken corpora is used to demonstrate how the spoken language is used in real life and in lots of different circumstances. The earliest spoken corpora made electronic in history was London-Lund Corpus of Spoken English (LLC) which stayed relatively the most widely used one for a long time. It captures approximately 500,000 words along with the speakers at University College, London were mainly faculty, staff, and students.

b. Specialized Corpora

Researcher creates the Specialized Corpora for the specific reason that is based on their study purposes. Specialized corpora might include certain textual genres. Specialized corpora may be narrow or wide in variety and are mostly built to resolve specific questions (Bennet, 2010, p. 13). The Michigan Corpus of Academic Spoken English (MICASE), which includes only spoken language from a university setting; the CHILDES Corpus that consists language used by children; and The International Corpus of Learner English (ICLE), are a few examples of the standardized corpora.

c. General Corpora

General corpora is meant to reflect the whole of a language or variety. It includes both spoken and written language, different forms of text etc. The Bank of English (BoE), the British National Corpus (BNC), the American National Corpus (ANC), the Corpus of Contemporary American English (COCA), and the International Corpus of English (ICE) are corpora consisting both written and spoken languages. Furthermore, there are also corpora that contains only written language or spoken-only language.

d. Parallel and Multilingual Corpora

The linguist collects this corpora not only consists texts in the English language but also two or more languages. They typically contain originals and their translations in different languages, or

related types of text. The idea of this type of corpora is to serve two or more different languages within the same types of text. It is commonly used for comparative linguistic studies and studies of translation. Parallel corpora generally contains the original form of a language and the translations, that is, it stores similar types of text with different languages.

e. Diachronic (historical) Corpora

The purpose of the diachronic (historical) corpora is to describe the use of an earlier or older phase of the language. It is used to compare the older and the modern texts. The linguists can use this sort of corpora to investigate how the language has changed over time. Examples of historical corpora containing texts from ancient English period: The Helsinki Corpora, and Lampeter Corpus of Early Modern English Tracts.

f. Text archives as Corpora

Text archives refer to databases of text. This is used to signify a collection of text that is being put together for a particular interest. Text archives which could be used as corpora are an online newspapers or CD-ROM, The Time Magazine Corpus, Oxford Text Archives (OTA), and Project Gutenberg.

g. Dictionaries as Corpora

Dictionary types that could be used as corpora are the electronic version if they constitute authentic examples as illustration. Examples

of corpora dictionaries are Dictionary of Old English (DOE), Middle English Dictionary (MED), and Oxford English Dictionary (OED).

h. The Web as Corpus

The web corpus emerges until the World Wide Web exists which can be used as sources for linguistic enquiries.

For a period of years, corpus-based analysis was applied to help develop and test language teaching and language learning vocabularies (Kaewphannagam, Broughton, & Soranasataporn, 2002). Thurstan and Candlin (1998) also posit the idea in order to deal with vocabulary selection and grading aspects, it is required to use a corpus-based study technique that allows teachers, course designers and material writers to select and grade the most frequently used items in order to respond directly to the learner 's objectives (Kaewphannagam, Broughton, & Soranasataporn, 2002). It is much easier to count the word frequency once analyzing a bunch of words by involvement of computer-based in research than using manual calculation. Corpus analysis, in other words, is a kind of language analysis that happens naturally by the use of corpora and computerized media. Hence, the analysis of the corpora better is conducted with the assistance of computer based programs, for instance to generate the frequency, corpus specialized software is used.

We can see how language is used in contemporary society by using the concept of the corpora. The researcher will consider the different ways in which languages are currently used when conducting a corpus analysis.

We also can consider these words and phrases as the most common in English by using corpus analysis (McCarthy, 2004, p. 1). The differences between spoken and written English, the differences between formal and informal words, how many words a learner needs to know to join daily conversation.

Corpora has been implemented in a broad range of fields, including analysis of translation, stylistics, grammar, and development of dictionaries (John, 1997, quoted in Laurence, 2004). In addition, the linguist can use corpus to study more information and get more precise frequency measurements (Lindquist, 2009, p. 5). Corpora analysis generally found out the results of study into concordance figures or words frequency list. The following are tools usually applied in corpus based analysis in order to discover the findings:

a. Concordance Tool or Concordancer

Concordance shows a list of how words and phrases usually occur in the context of a given text. Concordancer is sort of a language analysis search engine which scans through a corpus and lists the rate of every single words and phrases (Grigaliuniene, 2013, p. 37). To utilize a concordance, a word or phrase is entered and all instances of that word or phrase are found in the software. It is certainly faster to get the information in this way than reading through the full text (Lindquist, 2009, p. 5).

b. Word Frequency Lists

Corpus frequency list can be used to help classify words or phrases which are most common and least common. It refers to a list of all the words that appear in a corpus (Grigaliuniene, 2013, p. 43). Some of the key benefits of the frequency list device is that linguists can conveniently extract frequency data from vast volumes of text, which would be nearly impossible to manually accomplish (Lindquist, 2009, p. 8). Given the fact that frequency data cannot be used as a learner's only source of knowledge when selecting which words to know, it can be a good starting point.

c. Collocations

Collocation pertains to two or more words occurrence within a short space of each other in a text. Collocation is the statistical propensity of terms to co-occur (Bennet, 2010). It ensures that when one word is used, there is a strong statistical likelihood that there will be a similar word or words alongside it.

Regarding short stories, there are a lot of definitions of it, a short story is a full developed narrative which is shorter than a novel book and longer than a fable story. It typically takes only one single sitting to read. Short Story focuses on the incidents bigger or smaller and evokes strong feelings from its readers.

The short story is a concisely prose fictional type. The origins of the short story lie in the Middle Ages and Antiquity. The story, legend, and

fairy tale refer to the oldest forms of literary representations, "texts" that were transmitted mainly orally. The term "tale" (from "to tell") reflects this oral dimension inherent in short fiction, like the German "sage" (from "sagen"—"to speak") (M. Klarer, 2004: 13). One key aspect of the short story is that it is a short story (Cavanagh, 2010, p. 103). It can thus be distinguished from a novel, which is a long story.

Oberstein (1999: 15) mentioned that the definition of children literature is a category of books that existence of which absolutely depends on supposed relationships with particular reading audience: children. While short story define as prose fiction that is shorter in length than novel, so it can be concluded that child short story is prose fiction with concise form which is absolutely aimed to particular reading audience i.e. children.

In a short story analysis, there are lot of elements that can be considered might as well be applicable and can be used for when analyzing the corpus of the short story. A children short story was built from several elements that the researcher can look upon it when identifying the short stories. The following are the common elements in short story text:

a. Plot

Plot is the conceptual relationship of a text 's different thematic elements leading to a shift in the original situation as introduced at the narrative's outset (Klaler, 2004, p. 15). Gordon and Keuhner (1999: 1) said that the plot describes an event in a story as being deliberately planned by an author to achieve a desire result (Gordon & Kuehner,

1999, p. 1). And we may assume the plot is a sequence of acts and chronologically presented. But we need causality in order to make a good plot and connect each other between the incidents. There's little that happens without cause in narrative with a plot. So it will make a connection between the incidents in plot with the use of causality in plot. Plot traditionally grows out of a conflict. While conflict is a struggle internal or external between the main character and the opposing force. When a story involves the internal struggle, the main character is at odds with himself. Conversely, there could be an ongoing confrontation between the main character and either another entity, culture or natural nature, like destiny. In a plot there is an architecture or structure to illustrate the Fixing an accident (Gordon & Kuehner, 1999, p. 2). German critic Gustav Freytag drew up a diagram designing the plot structure.
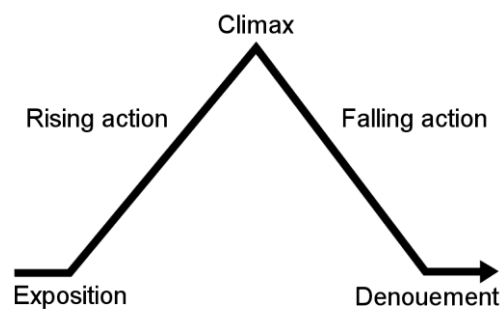

Figure 2. 1 Diagram of Plot Structure

Exposition is the explanatory information which the reader needs to understand the story situation. This is one of four basic forms of speech. Exposition sets the setting, the main characters, and

background information on what happened before the story started. Climax was defined as the point of the greatest conflict in the plot, the emotional high point, or the turning point. The conclusion of a plot also demands that the main character select some type of action which lead them to events called falling action. In short story falling action tends to be brief. It then usually lead to resolution or denouement.

b. Character

A person created for the work of fiction is a character of the literary word (Gordon and Keuhner, 1999, p. 95). The attributes in literature are graded to vary. First of all, it's round and flat. The round character is sufficient to surprise the reader, with three dimensional complex characters. Thanks to these qualities, there are other functions. The flat character, though, is one that cannot confuse the reader. These two-dimensional character will also be restored for one or two characteristics. The second are major and minor. In all literary works in particular, the narrator corresponds to the primary or central character. Generally speaking, this character is the most accurate and round character in the plot or any fictional work.

Moreover, the protagonist's opponent is then antagonist. The two characters then belong to the major character. The other major character is minor. The minor character may also be used as companion or a friend or two. The third character is active and static. Active character is one that changes in what's going on in a plot. The

static character stay unchanged, their character is same at the end of story as the beginning.

c. Setting

The setting in Klaler (2004: 25) is the place in which the action in a text is created, the historical and social environment. It is another aspect of prose fiction traditionally included. The setting can be broad, concrete, or very detailed and can usually serve as a background for a story (Gordon & Kuehner, 1999, p. 45).

d. Theme

Gordon and Keuhner (1999:1999) describe the theme as a perspective from an author or general reflection on the human existence or situation conveyed by character, computerization and imagery. Typically the subject is implied in the language. The concept is also more complex and open to diverse interpretations because it must be interpreted from the reader.

e. Point of View

The narrative viewpoint is known as point of view. Klaler (2004: 20) defines the perspective as how people, events and settings are presented in a text. The action of a text shall either be interpreted by an actual, unspecified narrator (omnistarist) or a person engaged in an event (first-person narration) or addressed without further explanation (figure narration situation). However, there are two main

views according to Gordon and Keuhner (1999:151), we should know: first person (I), and third person (he, she, they).

f. Style

Style is a characteristic way of saying a thing (Gordon & Kuehner, 1999, p. 243). Style reflects its nature or personality in literary work. But style refers to the author's own diction, imagery, syntax and variety choices.

Analyzing corpus of verbs frequently used in story was not easy task to conduct if the researcher doesn't know the fundamental theories building it as there are two different classes of verbs, auxiliary verbs (auxiliary) and ordinary verbs, in English. Auxiliaries is a type of verb used to help form a tense or an expression, and it is generally referred to as helping verbs. Auxiliary verbs play a significant role in shaping speech. The speaker's tense, aspect voice, mood, or intention is indicated along with the main verbs. Instead of repeating ordinary verbs, auxiliaries are used. Examples of Auxiliary verbs are is, am, are, was, do, have, can, must, etc. Ordinary verbs are used for conveying the full sense of the action, such as walk, write, read, prepare, see. (Thomson & Martinet, 1986). Verbs can be used to build the main part of predicate in a sentence.

Verbs in English are divided into three verb forms. The infinitive form always precedes by the word "to", for instance to read, to see, to write, etc. The base form has a same form with the infinitives verb but without word "to". The conjugated form is used to indicate the tenses or in other words it

indicates when the action takes place. Verbs have the following forms: base form (e.g. look), s-forms (looks), past form (looked), ing-form (looking) and past/passive participle (looking) (Eastwood, 2002, p. 75).

Furthermore, not only divided based on grammatical structure, verbs can also be classified according to their semantic meanings. Levin (1993) categorize verbs into at least 49 verb classes based on its semantic behavior that can be related to the syntactic behavior in sentence (Levin, 1993). . She first specifically distinguishes the various forms of syntactic action of the verb. Each of these forms are defined by one or more alternations (e.g. alternations describe passive forms; insertions and reflexive forms). Then according to these alternations, she suggests an overview of English verbs: each verb is correlated with the collection of alternations it undergoes. A preliminary investigation showed that there are sufficient similarities to allow the creation of classes between some aspects of verbs' semantics and their syntactic actions. Beth Levin then identified about 200 verb semantic classes from these observations, where, in each class, verbs share a certain number of alternations.

A collection of verb semantic groups is organized out of this grouping. Each of verb classes sometimes also has subclasses based on its lexical properties. For eg, we have the classes of putting verbs that include Put verbs, Funnel Verbs, Putting verbs in a defined direction, Pour verbs, Coil verbs, etc. Other verbs classes set are: Verbs of removing, Verbs of Carrying and Sending, Verbs of Throwing, Hold and Keep verbs, Verbs of

contact by impact, Image creation verbs, Verbs of creation and transformation, Verbs with predicative complements, Verbs of perception, Verbs of desire, Verbs of communication, Verbs of social interaction, etc. As can be observed, these classes only slightly overlap with the WordNet classification. Since the classification criteria are very distinct, this is not surprising. Apparently there are several way of groupings verbs aside from classification based on Levin Classes, such as:

VerbNet is a syntactic-semantic systematic summary of the Levin groups and is a lexicon consisting of approximately 5800 English verbs and group verbs (Kipper-Schuler, Dang, & Palmer, 2000) (Kipper-Schuler, 2005) in (Jayan & Kumar, 2018). It is a hierarchical, domain-independent, web broadcast verb lexicon. It consists of around 270 verb groups and is a domain independent lexicon. The same syntactic frames that have the same syntactic actions are shared by and VerbNetverb class entry. In several NLP applications, VerbNet is efficiently used.

A different approach is to use Frame Semantics to define the verbs. The FrameNet finds insemantic position assignment, comprehension of natural language, machine translation, textual entailment, extraction of knowledge etc. in its applications. At present, FrameNet has approximately 170,000 manually annotated sentences that provide forsemantic position labeling with a specific training data set (Baker, Fillmore, & Lowe, 1998, August). The terms are arranged in a hierarchical order with a reference to a particular definition in FrameNet data-set.

Another critical lexical resource for semantic role labeling (SRL) is PropBank (Palmer & Kingsbury, 2003). PropBank is a bank of proposals in which sentences and their statements are annotated with verbal propositions. This was suggested by Martha Palmer, et. al. It is similar to FrameNet, but differs in two remarkable ways (Dutta, 2011). It was used extensively for the CoNLL Mutual Mission 2004-05 SRL Task (in Jayan & Kumar, 2018).

In addition, the approach of Chafe deals primarily with the essence of the verb that the noun accompanies (Chafe, 1970). Next, the verb classification of Cook is a two parameter-based matrix system (Cook, 1979). The vertical parameter consists of four fundamental verb types: state verbs, process verbs, action verbs, and action-process verbs. The other criteria consists of the roles, namely agent and/or patient, experiencer, benefactor or octave (in Jayan & Kumar, 2018).

For further explanation of each verbs classes in Levin, is going to be listedto some points below (Levin in Zahara, 2018):

1. Verbs of putting, refer to putting an entity, such as place, sit, drop, pour, fill, etc.

2. Verbs of removing, relate to the removal of an entity from a location, such as eject, remove, clear, erase, etc.

3. Verbs of sending and carrying, relate to causing an entity to change location and describe the causation of accompanied motion, such as bring, carry, take, post, send, etc.

4. Verbs of exerting force: push/pull verbs, relate to the exertion of a force on an entity, such as press, pull, push, etc.

5. Verbs of change of position, such as feed, give, allocate, offer, etc.

6. Learn verbs, describe the acquisition of information, such as acquire, glean, learn, memorize, study, etc.

7. Hold and keep verbs. Hold verbs describe prolonged contact with an entity but do not describe a change of possession and location such as clasp, grasp, hold, handle, etc. Keep verbs relate to maintaining something at some location, such as keep.

8. Verbs of concealment, include verbs relating to keep something out of view, such as block, conceal, hide, etc.

9. Verbs of throwing, verbs relate causing ballistic motion by giving a force, such as slap, sling, throw, etc.

10. Verbs of contact by impact, refer to verbs of moving an entity to contact with another entity, such as beat, hit, kick, slam, etc.

11. Poke verbs, describe verbs of bringing a pointed object into contact with a surface and, in some instances, puncturing the surface, such as dig, poke, stick, etc.

12. Verbs of contact: touch verbs. It refers to pure verbs of contact, and they describe surface contact with no necessary implication that the contact came about through impact, such as graze, kiss, touch, etc.

13. Verbs of cutting, relate to a separation in material integrity, and also include some specification concerning the instrument or means used

to bring the result about the materials, such as chip, cut, hack, saw, etc.

14. Verbs of combining and attaching, refer to verbs that are related to combining and attaching, such as blend, combine, connect, join, etc.

15. Verbs of separating and disassembling, refer to all verbs that are related to separating or disassembling process, such as differentiate, break, etc.

16. Verbs of colouring, describe a process of changing the colour of an entity, usually by the application of some coating that covers the surface of the entity and changes its colour, such as colour, paint, varnish, etc.

17. Image creation verbs, contains verbs relating to the creation of images in the surfaces, such as paint, emboss, draw, etc.

18. Verbs of creation and transformation, contains verbs relate to create or transform an entity, such as build, make, work, etc.

19. Engender verbs, describe a causal relationship between the two arguments which are typically both abstract noun, such as beget, cause, generate, etc.

20. Calve verbs, describe the verbs relate to giving birth to the animal from which the verb takes its name, such as calve, cub, foal, kitten, etc.

21. Verbs with predicative complement, include verbs which have the predicative complement predicated of the immediately post-verbal noun phrase, such as adopt, nominate, proclaim, etc.

22. Verbs of perception, such as feel, hear, see, sense, etc.

23. Psychological state verbs, include verbs as the experiencer and the stimulus, or sometime theme, cause, object of emotion, or target of emotion, such as affect, amaze, confuse, envy, etc.

24. Verbs of desire, express the subject's desire, such as desire, fancy, want, need, etc.

25. Judgement verbs, relate to a judgement or opinion that someone may have in reaction to something, such as punish, persecute, praise, acclaim, etc.

26. Verbs of assessment, relate to verbs in making an assessment of something with respect to some attribute, such as analyze, assess, evaluate, etc.

27. Verbs of searching, such as dig, check, search, stalk, seek, etc.

28. Verbs of social interaction, relate to group activities that inherently involve more than one participant, such as agree, conflict, fight, meet, etc.

29. Verbs of communication, describe verbs relate to communication and the transfer of ideas, such as dictate, explain, say, call, chat, etc.

30. Verbs of sounds made by animals, such as chirp, snuffle, squeak, etc.

31. Verbs of ingesting, relate to the ingestion of food and drink, such as drink, eat, chew, breakfast, feed, gulp, etc.

32. Verbs involving the body, relate to verbs involving the bodily processes, such as sneeze, snore, breath, cough, blink, etc.

33. Verbs of grooming and bodily care, such as dress, groom, wash, wear, etc.

34. Verbs of killing, include all verb classes describe killing, such as execute, kill, shoot, poison, murder, etc.

35. Verbs of emission, involve the emission of a stimulus or substance that is particular to some entity, and consequently these verbs take a very limited range of subject, such as shine, sparkle, beep, smell, pour, etc.

36. Destroy verbs, relate to the total destruction of entities, such as waste, ruin, destroy, devastate, etc.

37. Change of state verbs, such as break, wrinkle, bake, boil, collapse, etc.

38. Lodge verbs, describe one's living situation, such as board, camp, live, settle, stay, etc.

39. Verbs of existence, relate to the existence of an entity at some location, such as exist, bloom, remain, fill, follow, etc.

40. Verbs of appearance, disappearance, and occurrence, describe the appearance, disappearance, and occurrence of an entity, such as appear, arise, expire, die, happen, occur, etc.

41. Verbs of bodily internal motion, describe the movements of the body, take animate subject, such as kick, flap, teeter, fidget, etc.

42. Verbs of assuming a position, describe the spatial configuration of the specific entity, such as bend, bow, kneel, sit, etc.

43. Verbs of motion, refer to verbs include a specification of direction of motion, such as advance, arrive, leave, move, fly, ride, etc.

44. Avoid verbs, verbs relate to avoidance, such as avoid, boycott, elude, etc.

45. Verbs of lingering and rushing, describe the shortening or the prolongation of the time course of an event, such as delay, hurry, rush, etc.

46. Measure verbs, relate to the value of some attribute of an entity with a relevant scale, such as read, cost, price, rate, charge, etc.

47. Aspectual verbs, describe the initiation, termination, or continuation of an activity, such as begin, continue, end, finish, repeat, etc.

48. Weekend verbs, relate to the spending a period of time at some location, such as holiday, honeymoon, sojourn, vacation, etc.

49. Weather verbs, describe different types of weather, such as blow, clear, freeze, snow, storm, thunder, etc.

B. Previous Studies

In this sub-section, the researcher presents some related studies that has been conducted earlier related to corpus analysis of frequently used

verbs in short stories. In this regards, the researcher only takes into account studies that concern with the topic related with this study. The following studies being reviewed are some studies which was conducted in regard to this field of study.

Study was undertaken in 2016 by Sun concerning on the analysis of vocabulary coverage in children short story entitled "The Magic Tree House". Study conducted based on the problem in second language learning of extensive reading. Teachers, students, and parents perceive extensive reading as impractical due to the lack of immediate measurable gain in test scores. Extensive reading is still not commonly practiced at EFL setting in Taiwan, according to Sun. Due to the lack of immediate measurable gain in test scores, teachers, students, and parents perceive extensive reading as impractical. Thus, reading a large amount of target language materials can help learners acquire new words and incidentally reinforce previously learned ones to overcome the difficulties. Sun uses a corpus analysis approach to examine vocabulary use within a single-author story sequences, its vocabulary repetition, and the incidental and intentional learning opportunities in it. The findings indicate that The Magic Three House series can support word list learning due its wide coverage of vocabulary from the three list, especially the Middle School Basic 1,200 list. To conclude, the results suggest that the more basic a vocabulary list is, the higher the repetition rate. This was shown by the data that among 48 books of the

series, they contain 91.4% of Basic 1,200 word, 36.4% of Advanced 800 word, and 36.4% of High School 4,700 words (Sun, 2016).

Furthermore, in the same year, Moon (2016) from the University of Washington conducted a study regarding to the similarities and differences in word use between two news reports corpora; The Time of London and Kenya's Daily Nation. Both of the news reports sections she used were taken from both news during 2013 which 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. The study comprises a total of 53.638 new articles, 1.992 from Daily Nation (759.720 words) and 51.646 articles from The Times (19.285.619 words). Based on the analysis, Moon agree 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. On the other hand, more space of international and global news and citizens were devoted in The Times (Moon, 2016).

Additionally, Yilmaz & Özdem in 2017 conducted a study centered on contrastive corpus analysis of reporting verbs used by native and non-native speakers. Their study aimed to investigate the most frequently used reporting verbs by native and non-native researchers in ELT and also investigate functional and positional differences in the use of these reporting

verbs. In resolving these concerns, they compiled two corpora of international journals in the field of language teaching and 160 ELT research articles published by peer-review. The ELT research articles were also divided into two based on the writer; 80 articles were written by native researchers and 80 articles written by non-native researchers. The results showed that non-native researchers overuse such reporting terms, such as revealed, indicated, and observed. There are also some variations between the two corpora in the use of reporting verbs in terms of syntactic patterning (Yilmaz & Özdem, 2017).

Study was also conducted by Oktavianti in 2019 aimed to examine the core modal verbs in English and their counterparts in Indonesian. The study exposes the plausible modality categories dominant among English and Indonesian speakers. Data was collected from several corpora; ARCHER and COCA for English language data and MCP and Wortschatz for Indonesian language data. She analyzed the data quantitatively using the normalized method of frequency and qualitatively using the method of translational identity. The analysis shows that core modal verbs expressing volition and prediction are most frequently found in English, with the modal verb will being the most frequently used modal verb. The contrastive analysis shows prediction and volition are important national states in terms of modality expression in both English and Indonesia (Oktavianti, 2019).

On the basis of some previous studies, corpora has been rapidly evolving over time. Corpora has a broad variety of work focused on several

preceding studies. Many work performed by the researcher in the corpus analysis, based on their purposes (McCarthy, 2004: 2) Corpora exist in several languages and is gathered for different purposes. Learner's dictionary, grammar reference materials, vocabulary learning materials, and later course books have all benefited from corporate information. However, only a few researchers attempt to examine the most commonly used verb in junior high school short story.

This presents study conducted by the researcher is different than the previous studies in terms of the types of corpora, tools of corpus analysis, and the focus items which were being analyzed in the corpus. In terms of tools of corpus, this study differs with the previous study as the researcher utilized Antconc 3.2.4 and TagAnt software in analyzing the short stories, while sun's study used a computerized tool developed by Visual studio #C which then the data was further analyzed into Excel file. Sun also used three wordlist from Ministry of Education in Taiwan which are Middle School Basic 1,200, Middle School Advanced 800, and High School 4,700, as guidelines for textbook developers to track. In addition, this study also contrary with the study from Moon which yielded the news from factiva metadata (news database) which then analyzed by using WordStat. Meanwhile, Yilmaz & Özdem study may have a bit resemblance with this study in the tools they used which is only Antconc 3.4.4. While Octavianti used normalized frequency method and translational identity method.

In terms of types of corpora, this study analyzed the corpus of children short story for junior high school students. This might resemble with sun's study that also analyzed corpus from children short story but it is only limited on the series of story within single author which is The Magic Tree House series (book 1 up to book 48). Meanwhile, Yilmaz and Özdem analyzed ELT articles and journals from native and non-native researchers. Also, Oktavianti used a combination of several corpora such as diachronic corpora and general corpora including ARCHER, COCA, MCP, Wortschatz.

The focus items in this study is to analyze the most frequently used verbs of children short stories for junior high school students. Meanwhile, the study from Moon concerned more generic onto comparison of part of speech (noun, adjective, verbs, pronoun) frequently used between two news report corpora. In addition, Yilmaz and Özdem focused their study on comparing specifically reporting verbs between native and non-native researchers' writings. Lastly, Oktavianti aimed to investigating specifically on frequency of modal verbs such as can, could, will, would, shall, should, may, might, and must with its equivalent words in Indonesian language.

Thus, the importance of this analysis would include the verb most commonly used for junior high school students in the short stories. Hence, the researcher intended in conducting a research entitled "**The Corpus Analysis of The Most Frequently Used Verb in the Children Short Stories for Junior High School Students.**"