CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter, the researcher presents the result of reviewing of related literature which covers the definition of corpus and corpus analysis, news report, verbs, and the previous study.

A. Corpus and Corpus Analysis

1. Definition of Corpus

The research which takes the idea of using corpus had been increased in this last five decades. Definitely, corpus, the plural forms is called as corpora, refers to the compilation of text, written or spoken, which are collected systematically. According to Oxford Advanced Learner's Dictionary (1995), the word corpus defines as a collection of written and/or spoken texts. In the present studies, corpus usually refer to a collection of machine-readable authentic text that chosen to characterize or represent a state or variety of a language (Grigaliuniene, 2013:09). In the same way, Linquist (2009) also agreed that corpus is related to electronic corpus. He stated that corpus is a collection of texts stored on some kind of digital medium to be used by linguists with the purpose of retrieving linguistic items for research or by lexicographers in making dictionaries (Lindquist, 2009:03).

According to McCarthy (2004: 1), he defines that corpus is a collection of texts, written or spoken, usually stored in a computer database. Renouf

(1987) and Sinclair (1991), cited by Ireneusz Kida (2013), also defines corpus in the similar way. Renouf (1987) argued that corpus refers to a collection of written or spoken texts which is stored and processed on computer for the purposes of linguistic research. While, Sinclair (1991) said that a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language. Hunston (2006) also agreed that corpus is collection of samples of naturally occurring language.

Regarding to the definitions above, it can be concluded that corpus is stored in a computer based system, contains entire forms of text, even they are written or spoken form. It can be said that computers has a significant roles in corpus. Computer based programs can be used in collecting, and storing the corpus data. The use of computer based programs makes a possibility in compiling a huge size of data quickly with a low budget. Moreover, computers can be used to manage the data of the corpus to conduct an analysis. Leech (1992) in McEnary and Gabrielatos (2006: 34) said that the use of computers gives us the ability to comprehend, and to account for the contents of corpora.

Besides, the text which used in corpus is compiled systematically. It means that the corpus is selected in certain structure and content based on the purposes in building the corpus. Hunston (2006) argued that usually the text are selected to represent a type of communication or a variety of language. For example, the first electronic based corpus was formed by Nelson Francis and Henry Kucera at Brown University in the early 1960s, and it is aimed to

represent the use of language in various text genres in present American English.

2. Types of Corpora

As corpora has been used in many linguistic research within various purposes, it makes the number of corpora is increasing rapidly. Each corpora has a different genre of the texts and different size of words number. Hunston (2006) said that corpora can be varying on its size, compiled in different purposes, and composed of text with different types. Corpora can be all homogeneous or all heterogeneous to a certain extent. Corpora which all homogeneous are composed of texts from one language or one register, etc. Whereas, corpora which all heterogeneous are composed from a number of different texts. Besides, corpora may be built in different size for specific reason. For example, a very huge corpora can be used to build a dictionary.

According on Lindquist (2009: 10) there are eight types of corpora of the English language:

a. Spoken corpora

According to McCarthy (2004: 1), spoken corpora contain transcripts of spoken language, such as transcripts of ordinary conversations recorded in people's homes and workplaces, or of phone calls, business meetings, radio broadcasts, or TV shows. Spoken corpora is used to represent how spoken language is used in real life and in many different contexts. Several corpora as the example of spoken corpora are the London-Lund

Corpus of Spoken English (LLC), Corpus of London Teenage Speech (COLT), Australian Corpus of English (ACE), Wellington Spoken Corpus (WSC), Longman Spoken American Corpus, and the Santa Barbara Corpus. London-Lund Corpus of Spoken English (LLC), and for a long time it remained the one most frequently used. It contains around 500,000 words and the speakers were mainly faculty, staff and students at University College, London.

b. General corpora

General corpora is aimed to represent a language or variety as a whole. It contains both spoken and written language, different text types, etc. Corpora that contain of both written and spoken language are The Bank of English (BoE), The British National Corpus (BNC). The American National Corpus (ANC), Corpus of Contemporary American English (COCA), and the International Corpus of English (ICE).

c. Specialised corpora

The Specialised corpora is created by the researcher for the specific reason based for their purposes of the study. Specialised corpora may contain certain genres of text. Bennett (2010) stated that specialised corpora can be large or small and are often created to answer very specific questions. The examples of the specialized corpora include The Michigan Corpus of Academic Spoken English (MICASE), which contains only spoken language from a university setting, the CHILDES

Corpus, contains language used by children, and The International Corpus of Learner English (ICLE).

d. Historical (diachronic) corpora

Historical corpora is aimed to represent the use of an earlier phase or older phase of language. It is used to compare the older texts with the modern ones. This kind of corpora can be used by the linguists to investigate how the language has change over time. The examples of historical corpora which contain texts from older period of English: The Helsinki Corpora, and Lampeter Corpus of Early Modern English Tracts.

e. Parallel and multilingual corpora

This type of corpora is aimed to represent two or more different languages within the same text types. It is usually used for comparative linguistic studies and translation studies. Commonly, parallel corpora contains of the original form of a language and the translations, in other words it stores a similar text types with different language.

f. Dictionaries as corpora

Types of dictionaries that might be used as corpora is the electronic version if it contains of authentic examples as illustration. The example of dictionaries corpora are Dictionary of Old English (DOE), the Middle English Dictionary (MED), and the Oxford English Dictionary (OED).

g. Text archives as corpora

Text archives refer to text databases. This is used to indicate a collections of text which are put together for certain interest. Text archives that can be used as corpora for instance newspaper online or on CD-ROM, the Time Magazine Corpus, the Oxford Text Archives (OTA), and Project Gutenberg.

h. The web as corpus

The web corpus appears when the existence of the World Wide Web that can be used as a sources for linguistic investigations.

3. Corpus Analysis

Kaewphannagam et.al. (2002), corpus-based analysis has been applied for a period of time to help develop and evaluate vocabulary for the teaching and learning of a language. He also cited that according to Thurstan and Candlin (1998) to deal with the aspects of vocabulary selection and grading, it is necessary to use the technique of a corpus-based study that permits teachers, course designers, and material writers to select and grade the most frequent items, so that it responds to the learners' objectives directly. In analysing a lots of words, it's much easier when computer based is involved in the study to count the word frequency than using manual calculation. In other words, corpus analysis is kind of analysis of language that occurs naturally by using corpora and computerized program. Thus, the analysis of the corpora is conducted with

the help of computer based programs, for example to generate the frequency, corpus specialized software is used.

By using the idea of corpora, we can see the way language is used in contemporary society, the way language's changed over the time, and the way language is used in different situations. In having a corpus analysis, the researcher can find the different ways of languages are actually used. In other words, corpus analysis can be said as an empirical study, because it analyses the actual use patterns in natural texts. McCarthy (2004:01) describes that by using corpus analysis we can find such as the most frequent words and phrase in English, the differences between spoken and written English, the most frequent tenses which people use, what preposition which follow particular verbs, the way people use words like can, may, and might, the differences between formal and informal words, the way how and why people use idiomatic expression, how many words which must be known by a learner to join everyday conversation, the number of different words used by native speakers in conversation, etc.

John (1997), cited in Laurence (2004), said that corpora have been applied in a wide range of areas, including translation studies, stylistics, grammar, and dictionary development. Lindquist (2009: 5) conveyed that corpus can be used by the linguist to investigate more material and get more exact calculations of frequencies. The analysis of corpora usually results the findings that are showed in one of two ways, as a concordance figures or as

word frequency list. To figure those result the tools of corpus analysis is used. The tools including:

a. Concordancer

Concordance shows a list of how words and phrases are generally occurring in context of the particular text. Concordance can be used to find the example of certain words are use in text quicker than searching the whole texts. Grigaliuniene (2013:37) describes that concordancer is a kind of search engine designed for language study. It looks through a corpus and lists every single occurrence of a word or phrase. As Sun and Wang (2003) describe in Laurence (2004) that corcondancers have been shown to be an effective aid in the acquisition of a second or foreign language, facilitating the learning vocabulary, collocations, grammar, and writing styles. The way how to use concordance is a word or a phrase is entered and then all the examples of that word or phrase will be searched by the software.

b. Frequency Lists

An information about frequency has a very significant role in language research and language learning. According to Grigaliuniene (2013:43), corpus-based frequency list is a list of all the words that occur in a corpus. In spite of frequency data cannot be used as the only source of information for a learner when

choosing which words to learn, it can be a good starting point when prioritizing which words that have to be learned first. (McCarthy, 2004: 10) also describes that frequency is the most basic tool for analysing the texts and contain list that tells us what words and phrases are used most often. Lindquist (2009: 8) added that one of the main advantages of frequency list tool is that linguists can easily get frequency data from large masses of text, which would be virtually impossible to achieve by hand. Frequency list can be sorted alphabetically or by total of the frequency that can be used to help to identify the most and the least common words or phrases used in a corpus. Thus, it can be used by the researcher to help in selecting the most frequent or the least frequent lexical items in language research.

c. Collocations

According to Sinclair (1991) in McEnery and Gabrielatos (2006:42) collocation refers to the occurrence of two or more words within a short space of each other in a text. According to Bennett (2010), collocation is the statistical tendency of words to co-occur. This means that when one word is used, there is a high statistical probability that a certain word or words will occur alongside of it. By using computerized corpora, we can study collocation by observing concordance lines to see which words occurs to the right

or to the left of the search word. Collocation also can be used to help us in understanding particular words used in a certain phrase.

B. News Report

Journalism is one of the most important part of life in this world. It is used to informed people about various events that happened in the community, state, country, and others. As stated by Mencher (2011: 56) that news is information which people can use to help them to make decisions about their lives, thus people need numerous information around them to enable to live well. These kind of information might spread to people through news media such as television, radio, magazines, newspaper, and many others. By using mass media the current and up-to-date information or ideas can be communicated to the public. News media is used to deliver the news items or news reports from the journalist to the readers or audiences in order to make the audience can comply their necessity of information. According to Gerot and Wignel (1998: 158), news item is a kind of text that used to inform readers, listeners, or viewers about events of the day which are considered newsworthy or important. While according to Deborah Potter (2006: 7), news can be found in one of three basic ways, that are naturally occurring events, like disasters and accidents; planned activities, like meetings and news conferences; and reporters' enterprise.

Regarding to how important news for people's daily life, the news that delivered to the public must be considered newsworthy to make the public

interested to read the information. There are several news values as the factor of making the news is newsworthy. According to Mencher (2011: 57), there are eight news values that must be noted in making news report, including:

1. Timeliness, refers to events that are immediate, or recent. News must be served directly after the occurrence to make it more attractive in public.

2. Impact, refers to news that are significant, important, and likely to affect many people. The more people who are affected by the events, the bigger the story.

3. Prominence, a news that deliver event involving well-known people or institution is more attractive than involving the less one.

4. Proximity, refers to news include events that are geographically or emotionally close to people, or usually known as localizing the news. People are interested in events and individuals that close to them.

5. Conflict, refers to strife, antagonism, and warfare have an attractive story for public.

6. The unusual, events that different from the expected of public and depart considerably from the experience of everyday life may attract the public attention, for example events or something that truly different, bizarre, strange, and wondrous in public.

7. Currency, refers to an event that currently happened will emerge as the subject of discussion and attention.

8. Necessity, means that the news writer has discovered something that he or she feels it is necessary to disclose to everyone. The news writer must consider the event to be something that should be exposed and known by public.

Mencher (2011: 30) also proposed that news story which delivered to the public must be:

1. Accurate, all information is verified before it is used, in other words a news have to provide accurate information.

2. Properly attribute, the reporter identifies all sources of information.

3. Complete, the story contains the specifics that illustrate, prove and document the main point of story.

4. Balanced and fair, all sides in a controversy are presented.

5. Objective, the writer does not inject his or her feelings or opinions.

6. Brief and focused, the news story gets to the point quickly and keeps to the point.

7. Well-written, the stories must be clear, timely, and interesting.

News report or news story is written differently from other texts. Usually other text, for example essay, is started with introduction then followed by the body and ended by conclusion. But news story usually written directly focused on the main facts or the climax from the events that is informed. This kind of structure is purposed to make the readers easy to gain the important points only in the first few seconds of reading. Therefore, most of news articles

might be written in short paragraphs. The generic structure of news item proposed by Gerot and Wignell (1998: 158) including

1. newsworthy event(s), in this part the event being informed is recounted in summary form;

2. background events, it presents us the elaboration of what happened, to whom, and in what circumstances it happened;

3. sources, this part shows the comments that given by the participants, witnesses, and authorities expert of the event.

While the significant lexicogrammatical features of news item are including short, telegraphic information about the story captured in headline, use of material processes to retell the event, use projecting verbal processed in sources stage, and focus on circumstances.

C. Verb

Verb according to Oxford Advanced Learner's Dictionary (1995: 1323) is a word or phrase that used to indicate an action, an event, or a state. Verb is used to form the main part of predicate of a sentence. According to Thomson & Martinet (1986: 73), there are two classes of verbs in English, they are the auxiliary verbs (auxiliaries) and ordinary verbs. Ordinary verbs usually known as main verb. Main verbs occur independently in a sentence and convey the complete meaning of the action, for example walk, write, read, cook, see, etc. Auxiliaries is type of verb used to help to form a tense or an expression, thus it usually known as helping verbs. It occurs along with the main verbs to indicate

tense, aspect voice, mood or intention of the speaker. Auxiliary verbs have an important role in forming conversation, because in short answers, agreements, disagreements with remarks, addition to remarks, etc., auxiliaries are used instead of repeating ordinary verbs. Auxiliary verbs include word is, am, are, was, do, have, can, must, etc.

Generally verbs in English divided into three verb forms, those are the infinitive form, the base form, and the conjugated form. Verbs in the infinitive form always precedes by the word *to*, for example to read, to see, to write, etc. Then, verb in the base form has a same form with the infinitive verb but without word *to*, for example read, see, write, etc. While in the conjugated form, there is a change on the word of verbs. It is used to indicate the tenses or in other words it indicates when the action takes place. It shows whether the action is occurred in the present, past, or future form. Verbs in conjugated form include went, cooked, will call, etc. while according to Eastwood (2002: 75), verbs have the following forms, such as a base form (e.g. *look*), an s-forms (*looks*), a past form (*looked*), an ing-form (*looking*), and past/passive participle (*looked*) and these forms also contain regular and irregular forms.

While, Levin (1993) divided English verbs into at least forty three verb classes based on its semantically meaning, they are:

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 analyse, 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.

D. Previous Study

Previous study is the result of the research from the researcher before some research related to the corpus analysis of the most frequently used verbs in the news article has been conducted by some researcher of the study. The previous study is a study of the use of verbs in research articles was conducted by Reimerink (2006). This Study was carried out from the perspective of genre.

It focused on the lexis analysis. The objective of this study was to examine the semantic categories of the verbs used in each section of the medical article, that are abstract, introduction, methods, results, and discussion. The Corpora contains 30 articles about breast cancer published over 1999-2001, which taken from Journal Citation Report web page (http://jcrweb.com/jcrr_selection.pl). The researcher in this study was using Wordsmith tools and the Brill POS tagger in conducting the analysis.

As the result of the study, the researcher found that the most used verbs in each section of article including verbs indicating Existence in introduction, Action verbs in Method section, Existence verb in Result, Existence verbs in Discussion section, and Cognition followed by Existence Action, Perception, Change and Possession verbs was within very similar percentage in Abstract.

The present study was different from the previous study in term of the corpus data being analysed. The corpus of the previous study are in the form of research articles. Whereas, this study was conducted to analyse corpus in the form of news articles which was taken from Fox News. The second difference is the focused items which analysed from the corpus. The previous study examined the semantic categories of the verbs used in each section of the medical article, that are abstract, introduction, methods, results, and discussion. While this present study focused in analysing which verbs are used most frequent in news article.