

CHAPTER II

LITERATURE REVIEW

In this chapter, the writer presents the theories that used in analyzing the corpus linguistics. In corpus linguistics we analyze the language aspects where it is about a set of procedures or methods. The language that is used is from an online news or online journalism from the year of 2010 to the present time. Corpus has a relation with adjective collocation, yet it can be changed from ages to ages.

2.1. Online Journalism

The amount information is being digitized and placed online, bringing with it the potential to transform the journalistic research process. The benefits of using online journalism as a research tool are:

- a. The range of access to sources – documents, data or news archives, and the people;
- b. The amount of accessing – million of documents, stories and contacts;
- c. The speed of access – the collection of the editorial content of hundreds of newspapers, magazine articles, and company reports, etc, that is updated daily.
- d. We can do everything we want with the information that we have got.
- e. The ability to tap into debates, discussions and expertise through email discussion lists or newsgroups where it keeps our electronic ear to the ground.

The benefits of online as a research tool are we can do what we could before to be more extensively and quickly. We could probably get much

information and available contacts online by traditional method. When researching a story we may have many different sources of information. These can be include a previous stories, original documents raw data press release or information from individuals (Ward, 2002).

2.2. Corpus Linguistics

Unlike sociolinguistics, psycholinguistics, or neurolinguistics, the study of corpus linguistics is not a branch of linguistics, since “corpus” does not tell us of what is studied rather a methodology that contain a large number of related methods which can be used by scholars, mostly the linguists with many different theoretical leanings. According to Lindquist, the writer of the book of *Corpus Linguistics and the Description of English*, he implied that by using a corpus, the researcher can investigate more material and get more exact calculation of frequencies (p. 1). He also added that corpus is mostly synonymous with electronic corpus, where a collection of several texts are stored on some kind of digital medium and used by many linguists to retrieve linguistic items for research or by lexicographers for dictionary-making process (p. 3).

A corpus can be described as a large collection of original texts or scripts that have been gathered in electronic form according to a specific set of criteria. The four important characteristics of a corpus according to Bowker (2003) are authentic, electronic, large, and specific criteria. These four characteristics make corpora different from other types of text collections and it will examine each of them in turn (p. 9).

There are two ways of getting the results of corpora, as a concordance or as frequency figures. Concordance is a list of all the contexts in which a word occurs in a certain text and the data are mostly presented in keyword-in-context (KWIC) concordances which shows one line of context and the keyword centered (p. 5). For instance, most of them take the first ten lines to reproduce. While a frequency figures shows the higher frequency data from

large masses of text that would be virtually impossible to achieve by hand.

Corpus linguistics and generative grammar have had such an uneasy relationship, and to find out the role of corpus analysis in linguistic theory, the aims of generative grammar and the three types of adequacy (observational, descriptive, and explanatory). Investigating these three types of adequacy reveals the source of the conflict between the generative grammarian and the corpus linguist: while the generative grammarian strives for explanatory adequacy (the highest level of adequacy, according to Chomsky), the corpus linguist aims for descriptive adequacy (a lower level of adequacy), and it is arguable whether explanatory adequacy is even achievable through corpus analysis.

2.3. Diachronic Corpus

Language has changed over time, which means all aspects of language change over time. Meanwhile, an adopted of a ‘synchronic’ perspective and described the language at once is possible to take a freeze-frame picture. Today’s linguistic phenomena sometimes helps explain why their structure is the way it is. This is taking a ‘diachronic’ view. For instance, in the course of our corpus-based studies in this book we have already seen that in COCA (Corpus of Contemporary American), the use of *but* rose between the early 1990s and the early 2000s, while the use of *however* went down, that the term *global warming* replaced *greenhouse effect* in *Time* magazine between the 1980s and the 2000s, that the share of *whom* out of all *who* and *whom* went down from 9.3% in the 1920s to 3.7% in the 2000s in the Time Corpus, that *get*-passives increased in the same corpus in the same period, and finally that gender-marked nouns like *fireman*, *poetess* and *actress* were being replaced by gender-neutral words over the same period, but at very different rates (Lindquist, 2009). In fact, the study in corpora is language change. There are two major ways to do this, the first is studying ‘change in real time’ where we need to have a diachronic corpus, a corpus that contains some texts from different time periods and where the texts have been marked up for year of production. Alternatively, we can use similar corpora from different time periods.

Meanwhile, according to Meyer (2004) who clarified the goal of diachronic corpus which to give an opportunity to map and compare various fields or paradigms in successive diachronic stages in the past which ease the further researcher to determine from the 'time-frame' of the various historical periods of English are fairly well defined even though the complications can still arise (p. 68).

2.4. Semantics

A language consists of both words and grammar where on grammar has two parts as morphology that deals with the structure of words and syntax that deals with the way how words are combined together. Meanwhile, the organization of meaning where it is underlying both words and grammar called a semantic. On the book of Corpus Linguistics and Description (Lindquist, p. 57) that is written by Hans Lindquist, Sinclair suggested and explained by Stubbs, there are following way of the wording that has been simplified:

1. COLLOCATION is the relation between a word and individual word-forms which co-occur frequently with it.
2. COLLIGATION is the relation between a word and grammatical categories which co-occur frequently with it.
3. SEMANTIC PREFERENCE is the relation between a word and grammatically related words in a field.
4. SEMANTIC PROSODY is the discourse function of the word that describes the speaker's communicative purpose.

Meanwhile, the limitation of this study was done on the adjective collocation. However, the writer only used a collocation theory towards adjective. There are two types of divided collocations based on Lindquist's theory.:

1. Collocation in a Window, where the type of collocation refers to words with other word of collocates and occurred between the keyword but it does not necessarily stand in a direct grammatical relationship with it. The window refers to the space to the left and right of the keyword that is included in the search of such collocations (p. 73) .

2. Adjacent Collocations, unlike the previous concepts which is a statistical phenomenon and only looked at in the collocates occur immediately to the right or to the left of the keyword, it rather closer to the real linguistic structure when we choose a word that in speaking or writing are influenced by the words that we have uttered or written, and by the words we are planning to speak or write next (p. 78).

The conception of semantics under Stanley (2004) means a sentence relative to a context that is derived by taking the contents from the semantic part of a certain sentence and relate it to that context and composing them with the composition rules that gathering the syntactic structure of a sentence. This concept leads a semantic value of a sentence is determined by the speaker intentions together with features context. His purpose explained, because of semantics, the source of our intuitions about the truth and falsity of utterances has a relation with various possible circumstances.

2.5. Adjective Collocation

On the book of Lexicology and Corpus Linguistics, Halliday (2004) explain that collocation is the habitual meaningful that co-occurrence of two or more words in close to each other and its collocates or collocates (p. 168). The writer took an adjective classification over the grammatical class over semantics type. According to Dixon (2005) there are several types that related with the grammatical class adjective in English as follows:

1. DIMENSION, e.g. big, great, short, thin, round, narrow, deep, etc.
2. PHYSICAL PROPERTY, e.g. hard, strong, clean, cool, heavy, sweet, fresh, etc.; also includes a CORPOREAL subtype, e.g. well, sick, ill, dead, absent, etc.
3. SPEED, e.g. quick (at), rapid, fast (at), sudden.
4. AGE, e.g. young, old, new, modern.
5. COLOUR e.g. white, black, brown, red, golden.

6. VALUE, e.g. good, bad, lovely, perfect, odd, strange, curious, necessary.
7. DIFFICULTY—easy, difficult, tough, hard, simple.
8. VOLITION, e.g. deliberate, accidental, purposeful.
9. QUALIFICATION with a number of subtypes:
 - a. DEFINITE, a factual qualification regarding an event, e.g. definite, probable, true, obvious;
 - b. USUAL, the speaker's opinion about how predictable some happening is, e.g. usual, normal, common;
 - c. POSSIBLE, expressing the speaker's opinion about an event, which is often some potential happening, e.g. possible, impossible;
 - d. LIKELY, another opinion, but tending to focus on the subject's potentiality to engineer some happening, e.g. likely, certain;
 - e. SURE, as for (d), but with a stronger focus on the subject's control, e.g. sure;
 - f. CORRECT e.g. correct, right, wrong, appropriate, sensible.
10. HUMAN PROPENSITY has a number of subtypes:
 - a. FOND, with a similar meaning to liking verbs, e.g. fond (taking preposition of);
 - b. ANGRY, describing an emotional reaction to some definite happening, e.g. angry (with/at/about), jealous (of), mad (about), sad (about);
 - c. HAPPY, an emotional response to some actual or potential happening, e.g. anxious, keen, happy, thankful, careful, sorry, glad (all taking about); proud, ashamed, afraid (all taking of);

- d. UNSURE, the speaker's assessment about some potential event, e.g. certain, sure, unsure (all taking of or about), curious (about); (e) eager, with meanings similar to wanting verbs, e.g. eager, ready, prepared (all taking for), willing;
 - e. CLEVER, referring to ability, or an attitude towards social relations with others, e.g. clever, adept, stupid; lucky; kind, cruel; generous (all taking at);^[L]_[SEP]
11. HONEST, judgment of some person or statement as fair and just, e.g. honest (about/in/at), frank (in); ^[L]_[SEP](h) busy, referring to involvement in activity, e.g. busy (at/with), occupied (with), preoccupied (with), lazy (over).^[L]_[SEP]
12. SIMILARITY, comparing two things, states or events, e.g. like, unlike (which are the only adjectives to be followed by an NP (Noun Phrase) with no preposition); similar (to), different (from), equal (to/with), identical (to), analogous (to), separate (from), independent (of), consistent (with). (Dixon, 2005)

2.4. Related Studies

This research has similar studies and the writer used two of the similar research. The first research has written by Pradevi (2016) with the title of Adjectival Collocation for [Islam] in COCA and COHA during the Three Periods in the United States of America (Civil Rights Movement, World Trade Center Attack, and Obama's Presidency). This research discussed about the collocation of the word of [Islam] during the three periods in the United States of America. It discussed about how the collocation of word [Islam] changed from period to period as mention before. Pradevi focused her research on the Islam lexical item and used the COCA and COHA as the source of the data.

The second related studied was The Adjectival Collocates of the Noun 'Technology' from 1940 to 2015: A Study on American Corpora done by Ghaisani (2017). Her research is about the collocation that appeared after the noun

of [technology] by the year of 1940 until 2015. The author also used the COCA and COHA to gather the data. However, she took the data source of adjectival collocation of [technology] with different genres as follows fiction, magazines, and newspaper.

There are differences and similarities between the two related studies and the writer's studies. The similarities are found that these research are using Concordances technique which according to Lindquist (2009) which it is the method to describe the data by took only the first ten number from the lists of a words occurs in a particular text that mostly presented in keyword-in-context (KWIC) to get a better guide on the contents although we can sorting the concordance lines in various ways. Both of the previous studies are similar with this research because they are only focusing on one lexical collocation.

The two of them are using COCA and COHA as the source of data unlike the writer's where the data has taken from NOW Corpus. Another differences are found on the time period between this study where the writer only take the two presidencies which are Obama's and Trump's presidency. Meanwhile, the two previous studies are mostly taken three periods as Pradevi's study where she took the data from the year of Civil Rights Movement, World Trade Center Attack, and Obama's Presidency and Ghaisani's took the four period from the year of 1940, the World War II, until the year of 2016, the 21st century or present time.