

## CHAPTER II

### LITERATURE REVIEW

#### 2.1. Research in Corpus Linguistics

In analysing language in text, there are some ways to find out the typical vocabularies that is used. Corpus (linguistics) is the way that can make an analysis in the linguistic text more accurately and simply. It has been incredibly useful for many researches over the last few decades. Corpus linguistics is a relatively recent branch of linguistics, made popular since the advent of personal computers in the 1990s.

Corpus linguistics itself “is the study of language data on a large scale – the computer-aided analysis of very extensive collections of transcribed utterances or written texts” (McEnery & Hardie, 2012). In addition, Baker, Hardie and McEnery in the book entitled ‘A Glossary of Corpus Linguistics’ (2006) defined corpus as follows:

In linguistics a corpus is a collection of texts (a ‘body’ of language) stored in an electronic database. Corpora are usually large bodies of machine-readable text containing thousands or millions of words. A corpus is different from an archive in that often (but not always) the texts have been selected so that they can be said to be representative of a particular language variety or genre, therefore acting as a standard reference.

Above all, the term corpus is actually the Latin for ‘body’; hence, it is right if the corpus may define as the body of the text. Corpus in modern linguistics, “in contrast to being simply any body of text, might more accurately be described as a finite-sized body of machine-readable text, sampled in order to

be maximally representative of the language variety under consideration” (McEnery & Wilson, 2001, p. 32).

A common criticism of corpus; *‘corpus linguistics is only a quantitative approach, just useful for identifying general patterns but not for any in-depth qualitative analysis’* is not entirely correct. Corpus linguistics actually needs both of quantitative and qualitative techniques as Biber et al. (1998 in Baker, 2010, p. 10) emphasize in Baker’s ‘Sociolinguistics and Corpus Linguistics’:

corpus-based research actually depends on both quantitative and qualitative techniques: Association patterns represent quantitative relations, measuring the extent to which features and variants are associated with contextual factors. However, functional (qualitative) interpretation is also an essential step in any corpus-based analysis.

It means that besides the quantitative technique that researchers got from corpus, qualitative still appears as a crucial technique to support the analysis. In fact, corpus linguistics “is not able to explain why something is the way it is, only tell us what it is.” (Bennet, 2012). To answer and examine the question of why, we may use intuition, in which we would have not been able to be applied in quantitative technique.

In order to carry out on corpus data there are some techniques or processes, such as; frequency, concordance, collocation, keywords, and dispersion (Baker, 2010, p. 19). Not all of the techniques are required for this study. The most used technique would be the frequency technique which lists all words appearing in a corpus and specifies for each word how many times it occurs in that corpus.

There are two approaches of corpus; corpus-based and corpus-driven linguistics. “Corpus-based studies typically use corpus data in order to explore a theory or hypothesis, typically one established in the current literature, in order to validate it, refute it or refine it.” (McEnery & Hardie, 2012, p. 6). Lexicographers used this corpus-based approach to provide a better understanding of English, such as to know a better use of phrases usually use. Meanwhile, “corpus-driven linguistics rejects the characterisation of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language.” (McEnery & Hardie, 2012, p. 6). Since this study wants to extract the generalization of the data, the corpus-driven is the better approach to be taken during the analysis.

## **2.2. Sociolinguistics and Corpus Linguistics**

The study of corpus linguistics can be connected to another discipline, such as sociolinguistics which discusses about the relation between language and social class. Wardhaugh (2006, p. 1) affirms that sociolinguistics is “the relationship between language and society” or “the various functions of language in society”. Sociolinguists are interested in explaining the different languages that are used in different social contexts. The different social context may produce a different language, especially because of their social background.

As mentioned previously, corpus needs both quantitative and qualitative techniques. Corpora are usually carefully constructed so that they can be said to be representative of a particular language or language variety (Baker, 2010, p. 6).

Furthermore, Baker (2010, p. 7) indicates that a corpus-assisted methodology could be helpful for sociolinguistics as in fact, “corpus linguists and sociolinguists already share a number of fundamental tenets of best practice when it comes to linguistic analysis”. Besides, both of sociolinguistics and corpus linguistics do need quantitative methodologies to make and construct comparisons of different populations, in which it focuses on differences and similarities, that can be facilitated with statistical tests (2010, p. 7).

Thereby, in this study corpus is being used significantly to identify or classify the users of those who use language based on their social class. Biber (1988 in Baker, 2010, p. 45) has attested the study of the sociolinguistics varieties, such as register, could be examined through corpus linguistics. His hypothesis was that this particular linguistic feature would be a good indicator of certain other linguistic features. For instance, if register contained a lot of past tense verbs, then a lot of third person pronouns were also liable. In this way, we could view groups of linguistic features as clustering together in various registers.

To accomplish the case, Biber carried out a multi dimensional factor analysis on the corpus data as a whole. The factor analysis identified five different ‘dimensions’ or sets of linguistic features. The dimensions are as follows:

1. Involved vs. informational production
2. Narrative vs. non-narrative discourse
3. Elaborated vs. situation-dependent reference
4. Overt expression of argumentation
5. Impersonal vs. non-impersonal style

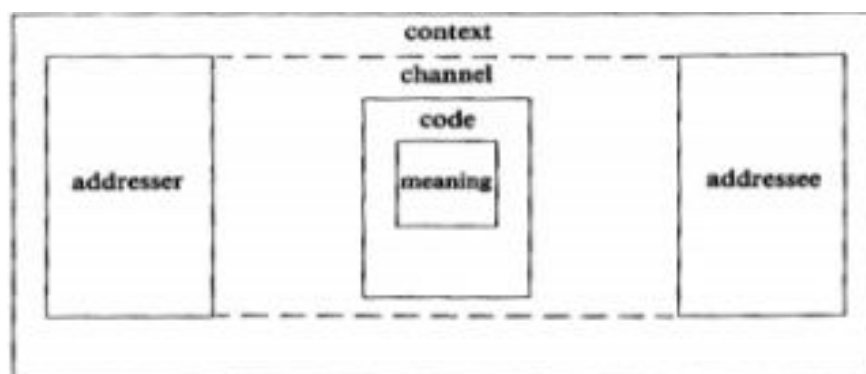
In each of the dimensions, the appearance (or non-appearance) of any feature would be a good predictor of whether the other features in that certain dimension would occur. In this case, connecting with register, each dimension could be viewed as a linear scale. By counting the presence of particular features with each dimension in the text, register could be placed at a particular point along that dimension at last. The study nevertheless, is one of the contributions of corpora toward applied linguistics.

Besides, there was also a study of linguistic variation and certain group or community (social class, in this case) in Norwich, England, which has been conducted by Trudgill (1974 in Wardhaugh, 2006, p. 149). He distinguishes five social classes; middle-middle class (MMC) lower-middle class (LMC), upper working class (UWC), middle-working class (MWC), and lower working class (LWC). Trudgill interviewed sixty informants which consisted of ten speakers from each of five electoral wards plus ten school-age children from two schools while the point scale are based on six categories; occupation, education, income, type of housing, locality, and father's occupation. The result shows that the using of the third person singular form of the present tense regular verb, the middle-middle class always uses *he goes* while the others use *he go*. It indicates that there is a sharp distinction between the middle-middle class groups and its lower groups. Moreover, what we can understand from Trudgill's study is it proves that the relation of (at least one of) linguistics variations and certain community or group (social class) actually exists in society.

### 2.3. The Language of Advertising

Advertisements, which is originally the word from Latin: 'advertere' (to turn towards), are the tool to advertise something. The communication through advertisement could be called advertising. Nowadays, advertising is everywhere in line with the development of the era. Advertising text usually occurs on newspapers, magazine, pamphlets and many others text media. The language of advertisements itself usually contains words which try to persuade and attract the target markets. Moreover, advertisements have an attractive power that is "able to manipulate the consumer; an invisible voice of advertisement advocates, encourages, asks, announces and deeply embeds into peoples' minds." (Lapsanská, 2006).

Since advertising is a special kind of communication, the language of advertisement is surely interesting to be discussed. Delivering the message only occurs if both of the people who post the advertisements and the potential customers have a communication process. Specifically, the process of communication in media can be represented in a picture designed by Leech (1974 in Vestergaard and Schroder, 1985, p. 15) as follows:



Furthermore, Leech gives the definition and explanation of the picture as follows:

the addresser is the copywriter, and the addressee is the reader, the meaning transmitted is about the product (more specifically, an attempt to make the reader buy the product), the code (in the case of press advertising) is language and some sort of visual code, the channel consists of printed publications, and the context will include such features as the reader's total situation (does he have the product already? Can he afford it? etc.), the publication in which the advertisement appears, and last but not least the knowledge that the text is and advert.

Embraced to this study about the classified advertisements in Jawa Pos, we can infer that the addresser is the people who post the advertisements, the addressee is the target consumers, the channel is Jawa Pos, the code is classified advertisements, and the meaning is the language that produced by the marketers to attract the target consumers.

In terms of the text advertisements, the 'meaning' that Leech means "may be descriptive information about the product, other text that serves the purpose of catching the readers's attention, as well as (typically) short phrases that act as a kind of slogan, and finally the name of the company and/or the name of the product." (Sells & Gonzales, 2002). The words or phrases that are used to attract potential consumers are very various. To give an indication of the result of this study, the writer would like to display the ten words or phrases that most frequently appear in a sample of television advertisement (Leech, 1966 in Sells & Gonzales, 2002):

❖ Ten most frequent adjectives:

1. New
2. Good / better / best

3. Free
4. Fresh
5. Delicious
6. Full
6. Sure
8. Clean
8. Wonderful
10. Special

❖ Ten most frequent verbs:

1. Make
2. Get
3. Give
4. Have
5. See
6. Buy
7. Come
8. Go
9. Know
10. Keep

Classified advertisements has their own group which do not make it appears as a single advertisements. Consequently, the competitiveness among the people who post the advertisements is inevitable. The only medium to attract the target customers is through their language of written text. And since the additional expense of adding a further line (in car classified advertisements in Jawa Pos in this case), the language that is used by the marketers “must also be attention grabbing, especially since obviously persuasive elements—such as catchy visuals—tend to be absent from these texts and very little can be done to induce a prospective public to read them.” (Bruthiaux, 1996, p. 24).

Thereby, the technique of wording has a significant role for the marketers to attract the customers. Briefly, the marketers should “stress the need for what



you have to offer, what it will do for readers, how they will benefit, what they can expect by using your product, how easy or more pleasant life will be for them.” (Classified Ad Secret).

#### **2.4. Related Studies**

There are some studies that are related to the writer's research. At the English Department of Airlangga University, the writer found two theses which also talk about advertisements. The first is the thesis entitled 'The Tendency of Choosing the Language Used in Vacancy Advertisements among the Young Job Seekers in Surabaya', which is written by Diah Aryanti (1997). Her thesis discusses about the young job seekers as the target reader of vacancy advertisements. Generally, there are two languages used in vacancy advertisements in Indonesia; Bahasa Indonesia (IA), English (EA). To make it specific, she divided the target of vacancy advertisements into two groups based on education background; senior high school graduates and university graduates.

Further, she wanted to find out the responses of young job seekers to the language used in vacancy advertisements, whether they have the same responses or not. She took sampling from particular respondents as well as an interview as the approach to accomplish the study. Later, it is found that senior high school graduates prefer to choose IA, while university graduates would rather to choose EA.

The similarities of her study with this study are both of the studies actually examine about advertisements and its relation with social class. Different from

this study, Arsyanti's examined about vacancy advertisements while this study examines about classified advertisements. Besides, her study also differs from this study with regard to the background of the social class. In this study, the social class is based on economic factor while the social class that was being researched by her was based on education background. In addition, as the writer wrote previously, Arsyanti took sampling and interview as the approach, contrast with this study which uses a corpus linguistic approach.

The second thesis which is related to this study is entitled 'Consumer's Preferences of the Language Style Used in the Simcard Product Advertisements Headlines', written by Agita Risma Nurhikmawati (2008). Concisely, the purpose of her study is to find out whether adolescent and adults have the same language style preference used in the simcard product advertisement headlines. In addition, to accomplish the study, she uses Poedjosoedarmo's theory on language style, Cook's theory on language of advertisement, and Holmes' theory on language choice especially in the part of social dimensions. Moreover, she uses quantitative approach; arranged the advertisements into table and then calculated into percentage and tested with chi-square. The result shows that the distribution of preference on the language style used has a different proportion for both categories of age. The adolescents prefer to use an informal style while adults prefer to use formal style.

Hence, that study is quite similar since it examines about advertisements and its relation with sociolinguistics. Unlike Nurhikmawati's study, the writer would like to examine about classified advertisements rather than headline

advertisements. Additionally, the involvement of sociolinguistics is because she used a social distance and formality scale whereas this study would focus more on social class. Nurhikmawati also used a quantitative approach in her research while this study uses a mixed method approach.