A Corpus-Assisted Discourse Analysis of Online Reviews

Rika Mutiara
Esa Unggul University
rikamutiara1205@gmail.com

Abstract

The purpose of the present study is to investigate how tourist attractions in Lombok are potrayed in English online reviews on TripAdvisor website. Keywords were investigated to find the prominent ideas in the reviews. The size of the study corpus is 95,694 words. This study used Brown corpus as a reference corpus to obtain the keywords. The collocations of the keywords i.e. 4 words to the left and right were scrutinized. The keywords and collocates were categorized based on USAS (UCREL Semantic Analysis System) category system to know the topics. The category that has the most members is movement, location, travel, and transport. Positive semantic prosody tends to occur when the topic is scenery in particular places such as *beach* and *waterfall*. Negative semantic prosody occurs in the collocations of some keywords that have negative senses such as *trash* and *rubbish*. Collocation of a neutral keyword to a negative collocate such as *dead coral* forms negative semantic prosody. In addition, collocations of two neutral words may create negative prosody. It can be seen by the occurrence of other words with negative senses in concordance lines, for instance, the occurrence of collocation of *sarong sellers* and *annoying*.

Keywords: keywords, collocation, semantic preference, semantic prosody

1. Introduction

The use of the internet as a mean to communicate becomes more popular. People from all over the world can share information easily. One of the ways to do it is by writing reviews of products or services in which they tell their experiences. Tripadvisor is a website that facilitate people to share their travel experiences. One's writing in this website is called a review. The reviews are classified towards the places or services they discuss. The products of such writings produce a new register. The study aims at finding how the tourist attractions in Lombok are potrayed by many reviewers.

To get the clear view of representation, corpus assisted-discourse study (CADS) was conducted. Applying the principles of corpus linguistics to analyze discourse is commonly done by textual method. In this method, language choice, meanings and patterns, phraseological elements were synthesized at the level of discourse (Hyland, 2009). Finally, it reveals what the discourse is about.

Some CADS have been conducted to investigate different types of texts. Koteyko (2012) used 79 compound words carbon as the starting points to analyze online blog discussion of climate change. Compound appears to shape thoughts in environmental issues, criticize government policy and show doubt. King (2015) split hapax legomena (lexis that occurs only once in the text) into some categories to see what the users of this chat service do with others.

The topic in this chat rooms is related to sex talk practices. It was found that the practices were for socializing. Cheng (2012) investigated western perceptions towards Hong Kong after 10 years of handover as represented in China and Western media. Particular lexis were chosen to know how Hong Kong is viewed in the Western and China corpora. In those cases, it can be seen that CADS gives contribution to potray social issues. Moreover, Reershemius (2012) investigated humor in academic presentation based on keyword analysis. It was used to build more interaction between the presenters and the audience.

Based on previous studies, it is necessary to set up a starting point to do the analysis. It could be particular lexis or hapax legomena. Another element in corpus linguistics is keyword. Keywords represent the issues in the texts. Keyness is defined as "a quality words may have in a given text or set of texts, suggesting that they are important, [that] they reflect what the text is really about" (Scott and Tribble 2006: 73). To obtain keywords, a study corpus needs to be compared with another corpus that is called a reference corpus. Statistical analysis reveals words that are keywords of the study corpus. Using keywords as a starting point in CADS is possible.

Another aspect of keywords such as collocation is a prominent element in CADS. To see how particular issues are represented, it can be started by observing the keywords and their collocates. A collocate is a word that co-occur with another word (keyword). Meaning of one word cannot be separated by words that frequently occur with them. Close concordance analysis reveal in what contexts usually the words occur and what meanings that are carried on. Analysing collocation reveals semantic preferences of the keywords and leads to semantic prosody. Semantic prosody can be considered to be neutral, positive, and negative senses of the texts.

There has not been a corpus-assisted discourse study of online review. The recent study tries to cover the semantic analysis of keywords and their collocations in online reviews. The recent study sought to answer the following research questions:

- What are the keywords?
- What are the topics of the reviews?
- How are the tourist attractions represented by the analysis of keywords and their collocates?

This study gives contribution on the understanding of discourse of online review as a new genre particularly those that discuss tourist attractions. In addition, it gives insights on how they were potrayed by the tourists to any parties that need such information.

2. Methodology

This study is a corpus-assisted discourse study. The size of the corpus is 95,694 words. It is a specialized corpus that consists of online reviews on Tripadvisor website written in 2015-2016 by native and non-native writers of English. The corpus was built manually by copying and pasting the texts into txt files. The texts in the corpus were run into Lextutor program to find keywords. This study used Brown corpus as a reference corpus because Brown corpus was made up by written texts. It makes both study corpus and reference corpus balance. The keywords that were analyzed are the first 50 keywords.

The collocates of the keywords i.e. 4 words to the left and right were scrutinized. The minimum frequency of the collocation is five. To get the collocations, Antconc software was used. Their collocations and keywords were categorized by using USAS semantic tagger. In addition, It gives information whether the word has positive or negative senses in the category of semantic classification. Each sense is classified to have scale ranging from 1 to 3. The categorization of keywords and the study of collocation and concordance lines show semantic preference and semantic prosody.

3. Findings and Discussion

3.1 Categorization of keywords and their collocates

This study focuses on the analysis of the first 50 keywords. The keywords occur in the family root. Therefore, a keyword that has a form as a countable noun might have singular and plural forms. A verb might occur in past, participle or progressive forms. An adjective might be in the forms of present participle or past participle. The keywords can be seen in the table below.

Table 1. Keywords

KEYWORDS				
waterfall	hike	trash	supermarket	annoy
trek	sarong	kilometer	souvenir	hindu
surf	mall	vendor	practice	vibe
yoga	metre	rubbish	swim	motorcycle
snorkel	disappoint	volcano	bungalow	Indonesia
bike	scooter	crater	campsite	guide
especial	motorbike	turtle	bamboo	meditate
amaze	centre	bracelet	coral	sunbathe
stun	refresh	colour	weave	jewel
clothe	donate	beach	hawking	equip

Keywords and their collocates were categorized based on USAS category system. The categories that exist are as follows:

Table 2. Categorization of keywords and their collocates

CATEGORIZATION	EXAMPLES
general and abstract term	donate, practice, had, equip, good, great, wonderful
the body and the individual	clothe, sarong, bracelet, weave
Emotion	stun, disappoint, annoy, enjoy
architecture, housing, and the home	centre, bungalow, restaurant
money and commerce in industry	mall, vendor, hawking, fee, paid, porters
entertainments, sports, and games	yoga, sunbathe, swim
life and living things	turtle, bamboo, coral

movement, location, travel, and transport	bike, motorcycle, swim, walk, surfing, beaches, village, temple
numbers and measurement	far, close, metre, kilometer
substances, materials, objects, and equipment	stunning, amazing, rubbish, trash, dump, water,
language and communication	said, explain, ask, recommended
social actions, states, and process	vendor, hindu, guide, visit, tourist, friendly
time	sunset, day, good, long
world and environment	waterfall, crater, hill, sandy
psychological actions, states, and process	view, experience, scenery, mediate
names and grammar	gilli, Lombok, senggigi, kuta I, we, you, Indonesia

3.2 Concordance analysis of keywords and collocation in each category

The top three categories that have the most members are movement, location, travel, and transport; general and abstract term; and substances, materials, objects, and equipment.

3.2.1 General and abstract terms

As keywords, most words were used to narrate what people do and describe the objects. Some words such as *donate*, *refresh*, and *practice* show what people did. People here might be the writers and their groups. Therefore, they told their experiences as can be seen in the following concordance lines:

Picture 1.
Concordance lines of *donate*

The writers also told possible actions that the readers might do based on their experience as in concordance lines 2 and 3. Based on USAS category system, *donate* was classified as a word with negative sense. When *donate* appears usually there is a bad condition happen that makes a donation need to be done.

3.2.2 Category of body and individual

Two items i.e. *sarong* and *weave* that become cultural characteristics of Lombok occur as keywords. *Sarong* often appears in the context of selling process. It commonly collocates with *sellers* and *but* as can be seen in the following concordance lines.

sistent itinerant sarong and bracelet sellers, but not overtly com

y a coconut or a sarong every 5 minutes but it is a really beautifu

are a couple of sarong sellers but not too bad. A gorgeous beac

n" as traditional sarong. But nice people indeed very open and fa

by bracelet- and sarong sellers. Annoying. But still better than Ku

Picture 2. Concordance lines of collocation of *sarong* and *seller*

In the concordance lines 3 and 5, it can be seen that the existence of sarong sellers gives particular feeling such as *not too bad* and *annoying*. Even though it gave negative feeling to the person, they still considered it is acceptable as can be seen in the line 3.

Weave tends to occur with the phrase *how to*. The writers told the activity of making *weave* as can be seen in the following concordance lines.

, etc. You can also learn how to weave with their traditional way

ition. You can also learn how to weave "tenun" and how to spin t

Picture 3. Concordance lines of *how to weave*

3.2.3 Emotion

Members in this category carry positive and negative senses. The analysis of USAS category system shows that *enjoy* have positive semantic sense. When *enjoy* collocates with *place*, they often occur in the phrase *place to enjoy* as in the following concordance lines.

nection and very strategic place to enjoy amazing beaches, mou

gigi is worth it to visit this place to enjoy the sunset. It is around

t many people. It's a nice place to relax and enjoy the scenery -

Picture 4. Concordance lines of *place to enjoy*

The occurrence of *enjoy* and *place* shows the writers point particular places where they enjoy the objects. *Disappoint* and *annoy* are keywords that show negative sense.

3.2.4 Architecture, housing, and the home

Campsite occurs in neutral sense but *campsites* appears in negative sense. They reported bad conditions as follows:

ainy season - the campsites are very basic - often without toilets

e the waste at the campsites as the rubbish are everywhere. It is I

Picture 5. Concordance lines of *campsites*

3.2.5 Money and commercial industry

Both traditional and modern commercial industry occur as *mall*, *supermarket*, *hawker*, and *vendor*. When the writers wrote about vendors, *vendors* occurs in neutral senses such as in

ere warned about aggressive vendors by our guide & true to his

nd the Senggigi coast. Street vendors will try to sell you things, I