การศึกษาทัศนคติต่อเหตุการณ์เยาวชนติดถ้ำโดยใช้ภาษาศาสตร์คลังข้อมูลเป็นฐาน

ปิยะพงษ์ เลาศรีรัตนชัย[1]
ศิริพร ปัญญาเมธีกุล[2]

(**Received:** *July 8, 2019;* **Revised:** *October 7, 2019;* **Accepted:** *November 20, 2019*)

บทคัดย่อ

เหตุการณ์เกี่ยวเยาวชนติดถ้ำเมื่อเร็วๆ นี้เป็นหนึ่งในเหตุการณ์ที่ผู้คนรู้จักกันอย่างดี และไม่ใช่แค่เหตุการณ์เพียงอย่างเดียวที่ทำให้ผู้คนรู้จักเหตุการณ์นี้ แต่ยังเป็นเพราะความสำเร็จในการช่วยเหลือเยาวชนอันเนื่องมาจากความร่วมมือ ร่วมใจของผู้คนจากทั่วโลกอีกด้วย วัตถุประสงค์ของงานวิจัยชิ้นนี้คือ เพื่อวิเคราะห์ทัศนคติของผู้ใช้งานทวิตเตอร์ผ่านภาษาที่ใช้ในการแสดงความคิดเห็น คลังข้อมูลภาษาประกอบด้วยคำจำนวน 16,621 คำซึ่งเก็บรวบรวมจากภาษาที่ใช้ในการแสดงความคิดเห็นในข่าวจากบัญชีทางการของสำนักข่าวบีบีซี (BBC) และซีเอ็นเอ็น (CNN) ในทวิตเตอร์และวิเคราะห์โดยพิจารณาจากความถี่ของคำ การวิเคราะห์คำสำคัญ การกระจายตัว โปรแกรม Linguistic Inquiry and Word Count (LIWC) และ Semantic Tagger ผลการวิจัยแสดงให้เห็นว่าวลีที่ผู้ใช้ทวิตเตอร์กล่าวถึงมากที่สุดคือ “the boys” นอกจากนี้ เมื่อพิจารณากราฟการกระจายตัวพบว่าวลี “the boys” นั้นกระจายตัวอยู่ทั่วคลังข้อมูลภาษามากที่สุดเช่นกัน แสดงให้เห็นว่า “the boys” ซึ่งหมายถึงเยาวชนติดถ้ำเป็นประเด็นที่ผู้ใช้ทวิตเตอร์เป็นห่วงและให้ความสนใจมากที่สุด นอกจากนี้ ผลการวิจัยที่สอดคล้องกันของ LIWC และ Semantic Tagger แสดงให้เห็นว่า ผู้ใช้ทวิตเตอร์มีทัศนคติด้านบวกต่อเหตุการณ์ครั้งนี้ และสรุปได้ว่า ไม่ว่าจะเกิดเหตุการณ์ใดขึ้น ผู้คนจากทั่วโลกยังคงสนับสนุนช่วยเหลือกันและกันเสมอ

คำสำคัญ: การศึกษาโดยใช้คลังข้อมูลภาษาเป็นฐาน ทัศนคติ เยาวชนติดถ้ำ ทวิตเตอร์

---

[1] ผู้ช่วยศาสตราจารย์, โครงการจัดตั้งสายอุตสาหกรรมบริการและนวัตกรรมภาษา คณะศิลปศาสตร์และวิทยาศาสตร์ กำแพงแสน, มหาวิทยาลัยเกษตรศาสตร์ Email: Pukan_walker@hotmail.com

[2] รองศาสตราจารย์ ดร., สาขาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

A Corpus-Based Study of Attitudes towards the Incident of the Thai Cave Boys

Piyapong Laosrirattanachai[3]
Siriporn Panyametheekul[4]

Abstract

The recent incident regarding the Thai boys trapped in a cave is one that is well known. Not only the incident itself made it well-known but also the successful outcome due to cooperation of people worldwide. The purpose of this study was to analyse the attitudes of Twitter users through the language used in their comments. The corpus containing 16,621 tokens was compiled by collecting the language used in the comments on news of the Thai cave boys from the BBC and CNN official accounts on Twitter.com and analysed using word frequency, keyword analysis, dispersion, Linguistic Inquiry and Word Count (LIWC), and Semantic Tagger. The results showed that the matter that the Twitter users mentioned most was "the boys". Furthermore, when considering the dispersion plots, the words "boys" occurred most across the corpus. This showed that "the boys" was the most interesting issue and also most concerned by Twitter users. The correlation results of LIWC and Semantic Tagger showed that the users had positive attitudes towards this incident. Whatever happens, people from different parts of the world still support each other.

Keywords: Corpus-based study, Attitude, Thai cave boys, Twitter

[3] Asst. Prof. of English for Service Industry Program, Department of Service Industry and Language Innovation, Faculty of Liberal Arts and Science, Kasetsart University, Nakhonpathom, Thailand.

[4] Assoc. Prof. Dr. of Linguistics Program, Faculty of Humanities, Srinakharinwirot University, Bangkok, Thailand.

Introduction

On 23rd June 2018, the news about twelve Thai boys and one man being lost was widespread throughout Thailand. Many organisations, groups of people, and individuals from both within Thailand and other nations offered their support in this incident. This was one of the greatest events in recent memory when many countries worldwide cooperated with each other. These Thai boys are all football players. After finishing playing football, they surveyed the cave at Tham Luang-Khun Nam Nang Non Forest Park, Chiang Rai Province. Unfortunately, after getting inside the cave, the rain was so heavy that the water level rose and was too high and blocked the entrance of the cave. This made it impossible for them to get out of the cave so they decided to escape the flood and went further into the cave. The incident lasted eighteen days, from 23rd June to 10th July 2018 when all members were rescued. During the incident, the news about the boys went viral and was broadcast on many channels—television, radio, newspaper, and online media. On virtual communities like Facebook, websites, Youtube, or Twitter, people always provided their comments on the latest news which can subsequently be used as a source for considering their attitudes.

According to Gardner (2002) and Baker (2006), research on attitudes towards language has been of interest since the 1920s. Conducting research on attitudes towards language involves interdisciplinary approaches between linguistics and socio-psychology. One of the most interesting research topics is studying people's attitudes through language used on online media. For example, the incident of the Thai cave boys news on Twitter, the high frequency word like "pray" might show that people were concerned about the boy and prayed for their safety. Studying the communication between users through computers, especially on social networking sites has been of considerable interest and is defined as "computer-mediated communication" or CMC (Herring, Stein &

Virtanen, 2013). Twitter is one of the most famous and popular social networking sites with about 335 million users (Clement, 2019). BBC and CNN, the well-known news agency of the United Kingdom and the United State of America respectively, also have their official accounts on Twitter. They launch a variety of news worldwide every day with the provision of space for Twitter users to like, retweet tweets, and leave comments.

The study aimed to explore the attitudes of the Twitter users towards the incident. To achieve the goal, this study investigated the attitudes of Twitter users towards the incident of the Thai cave boys through the users' comments and the topic in which they mentioned most. Corpus linguistics analysis was used to accomplish the purpose of the study. The data were collected from 10 posts from BBC and 10 posts from CNN official accounts reporting the Thai cave boys news during the period of the incident. Therefore, the corpora were self-constructed by collecting the comments left in 20 news posts from Twitter.

Research Objectives

The study aimed to investigate the attitudes of people, especially Twitter users, towards the Thai cave boys incident. To accomplish the purpose, the study was conducted to answer the following purposes.

1. To explore the topics in which Twitter users mostly pay attention to
2. To determine the attitudes of Twitter users towards the incident of the Thai cave boys

Literature Review

1. Corpus Linguistics

One of the misunderstandings about corpus linguistics is that people view it as a linguistics' branch or theory without being aware that it is not a branch of linguistics but a method of analysis in the linguistics field (Kennedy,

1998; Lindquist, 2009). Corpus linguistics is data-driven. Hence, data are very important in conducting corpus linguistics research. An advantage of using the data in any corpora is that the data are authentic rather than being language composed by researchers. Even though corpus linguistics is a powerful method for the analysis of language data, it has some limitations. With regard to using it to analyse media commentary, most programs are not capable of analysing emoticons, punctuation, stickers, or unconventional spelling. In this study, AntWordProfiler (Anthony, 2014) was used for lexical profiling, AntConc (Anthony, 2018) for frequency and dispersion, Key-BNC (Graham, 2014) for keyword analysis, and Semantic Tagger (Rayson, Archer, Piao & McEnery, 2004) and LIWC (Pennebaker, Booth, & Francis, 2007) for language attitude analysis. In this study, we mainly focused on words mostly used in comments and studied them using corpus linguistic approach. Any words often used in the corpus would be considered. These words lead to attitudes toward the incident. The texts in the corpus then was brought to support the analysis.

2. Frequency

Many scholars have agreed that the frequencies of occurrence are the best foundation for the phenomenon we investigate. Frequency, in linguistics, is one of the most impact concepts to study the ways of using language (Lindquist, 2009; Gries, 2010). In this study, the tokens in the corpus were calculated to determine the frequency of the words. The word frequency reflects the interest of Twitter users.

3. Lexical Profiling

Lexical profiling refers to categorising words. The categories are generated and we can make use of them as the base to allocate words. Examples of categories are: the General Service List (GSL) (West, 1953) —the

English words generally used in everyday life; the Academic Word List (AWL) (Coxhead, 2000) —the English words mostly used in academia; and Nation's Functional Word List (FWL) (retrieved from www.victoria.ac.nz/lals/about/staff/paul-nation) the function words used in English. One of the benefits of lexical profiling is that we can get rid of unwanted words from the corpus, leaving only the needed words. In this study, lexical profiling was used to find the words contained in the compiled corpus and those that also appeared in the FWL. Any words appeared in the FWL were ignored in the analysis, the remaining content words were analysed in the next steps.

4. Dispersion

According to Gries (2010), dispersion helps us to see the overall distribution of the focused words or phrases across a corpus. In addition, if any words or phrases appear across the corpus, it confirms that the focused words or phrases are not occurring by chance but they are significant and are relevant to the topics of the corpus. In this study, keywords were considered for their dispersion across the corpus, and dispersion plots were presented to support the data.

5. Keyword Analysis

Keyword analysis is the method of identifying the distinctive words of each corpus and is done by comparing the language in the target corpus with one or more corpora. The log-likelihood (LL) statistics is used to calculate the keywords. Keywords can be the signals showing that often-used words or phrases significantly imply some things such as the ways of talking or attitudes (Stubb, 1996). In the current study, keyword analysis was used by comparing the words in the corpus made from the CNN and BBC News' comments with the British National Corpus (BNC) using Key-BNC (Graham, 2014). The advantage

of using Key-BNC is that the program is user-friendly and designed to facilitate a researcher to calculate an LL value without approaching the BBC website.

6. Approaches to Studying Attitudes through Language

Studying attitudes towards language use has been of interest since the 1920s (Gardner, 2002; Baker, 2006). An attitude is a mental state expressed by assessing or evaluating something as good or bad, like or dislike, agree or disagree, etc. (Eagly & Chaiken, 1993). The attitude is subjective and varies from one item to another. Since it is a mental state, observing it only visually is rarely possible (Allport, 1935). According to Garrett (2010) and Ryan, Giles, & Hewstone (1988) (cited by Ivković, 2013), there are three main ways to conduct research on language attitudes.

6.1 Direct measure – the most basic and easiest way of data elicitation. The respondents are asked to rate the language use by filling in survey or questionnaire forms. However, the data elicited from this approach may be biased because people prefer positioning themselves with a good image socially. There is a high chance of obtaining a fake evaluation.

6.2 Indirect measure – this approach reduces the bias of response since respondents do not know what they are rating. The elicited data are more realistic and really come from the respondent's attitude. For this approach, "interpretation of the respondent" is the main source of data. For example, a picture might be shown to a respondent. A response from the respondent, which would be the interpretation of the respondent, then would be analysed to find the attitude. Apart from the interpretation, physiological responses like blinking, heart rate or blood pressure could also be used as one of the indirect measurement methods (Antonak & Livneh, 1995).

6.3 Societal treatment of language varieties – this approach suits the current study the most since it elicits the data from existing texts, which in this case are commentaries on Twitter.

In this study, the societal treatment of language varieties method was used since the researchers gathered the data from the commentaries on Twitter and analysed them using various steps.

Research Methodology

1. Corpus Construction

The corpus was compiled by collecting the language used in the comments on news of the Thai cave boys from the BBC and CNN official accounts on Twitter.com. Since news of any topics normally does not last long, the Thai cave boys news was presented about 20 times on the BBC and CNN official accounts each. Therefore, the corpus comprised the comments from 10 BBC news posts and 10 CNN news posts. The corpus is somewhat small with the total corpus size being 16,621 tokens.

2. Research Instruments

2.1 AntWordProfiler

AntWordProfiler (Anthony, 2014) facilitates a user by lexical profiling words in any corpus and showing the shared words between the target word list and the reference word list. In this study, the Functional Word List (FWL) was used as a reference word list. Any words from the corpus that appeared in the FWL were ignored, leaving the remaining words to be analysed. Most of the remaining words were content words which could show Twitter users' attitudes better than the functional words which mostly appeared to make sentences to be grammatical.

2.2 AntConc

Similar to AntWordProfiler, AntConc was created and developed by Anthony (2018). AntConc is capable of operating multifunctionally as it can calculate the frequency, show concordance, illustrate dispersion plots, etc. Thus, AntConc was used to calculate the word frequency and illustrate the dispersion plots in this study.

2.3 Key-BNC

Key-BNC (Graham, 2014) is a program used to calculate the "keyness" of the words contained in the corpus. The log-likelihood (LL) statistic is used in this program. Key-BNC helps the user to compare the word frequency of the target corpus with the word frequency of the reference corpus, (British National Corpus here), without accessing the website.

2.4 LIWC

Words expressed by people can inform us how they think, feel, or experience the event in the world (Tausczik & Pennebaker, 2010). The Linguistic Inquiry and Word Count, or LIWC, (Pennebaker, Booth, & Francis, 2007) is an online program used to analyse computerised text. The program itself can analyse the whole text and identify the proportion of positive and negative attitudes across the whole text. In this study, LIWC was used to analyse the positive and negative attitudes through the keywords in the comments on the Thai cave boys news on Twitter.com.

2.5 Semantic Tagger

Semantic Tagger (Rayson, Archer, Piao & McEnery, 2004) is the annotation system that annotates words depending on their semantic field. There are 21 major semantic fields with the provision of synonyms, antonyms, hypernyms and hyponyms (see http://ucrel.lancs.ac.uk/usas/ for more

information). The annotated words are tagged with the abbreviation representing each field. Furthermore, a positive or negative value was added to the tags using +/- symbols. Since the main focus of this study was to identify the attitudes, the semantic field results were ignored. The attitudes were analysed by considering the +/- symbols attached to the tags.

3. Data Analysis

In this study, the data were analysed using the following steps.

1. AntConc was used to calculate the frequency of the words contained in the corpus. The list of word frequencies was sorted from the most frequent to the least frequent.

2. Key-BNC was used to calculate the keyness of the word.

3. AntWordProfiler was used to create a lexical profile of the function words. The function words were then removed from both the word frequency list and the keyword list. The calculated keywords were then compared to the word frequency result to identify any correlation.

4. The number of words combined between the 10 most frequent words and the first 10 words with the highest keyness (function words were ignored here) were considered for concordance using AntConc through presentation as dispersion plots.

5. The attitudes of the commenters were analysed through the combination of the 50 most frequent words and the first 50 words with the highest keyness using LIWC and Semantic Tagger. The researchers used two programs here to increase the reliability, that is, to see if the results of the two tools supported each other. If they did support each other, then the results were reliable.

Findings

After complying the corpus, the data was analysed follow the following steps.

1. Frequency

Frequency of the words was analysed first. Table 1 shows the word frequency results using AntConc.

Table 1: The first fifty highest frequent words in the corpus

Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq
the	753	are	187	s	136	have	80	coach	64
to	442	you	182	be	131	On	78	who	63
and	405	this	179	out	125	these	78	thank	62
a	274	i	178	their	124	prayers	77	world	62
for	274	boys	172	rescue	108	cave	76	get	60
in	253	it	166	so	101	can	69	his	59
of	246	that	164	news	93	We	69	kids	59
all	238	them	155	not	86	good	68	bless	56
they	220	god	148	with	84	from	66	there	56
is	201	x	141	t	83	will	65	what	55

From table 1, most of the words are function words. However, there are some content words like “boys” with a frequency of 172 occurrences, “gods” with 148 occurrences, and “rescue” with 108 occurrences. Some of the function words like “all”, “they” and “out” are also interesting since this could have implied that commenters hoped that “all of them” (the boys) would finally come “out” of the cave. Apart from referring to the boys, “all” was also referred to the rescue team. See the following extract from the corpus.

Thread 1: "They're coming home...they're coming home...the Thai boys football team that got stuck in a cave are coming home"....excellent news that *all* rescued. Rescue crew are *all* heroes.

2. Keyword Analysis

Key-BNC was used as an instrument to identify the keyness of the words in the corpus. The results are shown in Table 2.

Table 2: First fifty words with the highest log-likelihood (LL) values

Word	LL	Word	LL	Word	LL	Word	LL	Word	LL
boys	1313.02	news	494.12	rescued	276.18	sad	180.44	involved	117.18
rescue	999.91	coach	472.79	diver	272.23	everyone	147.73	world	116.61
prayers	787.09	thai	467.27	rip	232.94	awesome	145.71	are	115.42
god	768.76	heroes	460.29	brave	227.90	out	145.58	save	115.40
cave	739.31	kids	405.80	thaicaverescue	226.37	amazing	144.18	caves	114.42
bless	604.86	all	372.67	pray	220.61	safely	139.92	children	111.32
praying	579.14	thank	308.46	thailand	217.02	oxygen	134.08	great	110.93
divers	564.66	hero	290.46	hope	215.26	let's	121.89	story	110.65
cnn	545.42	safe	288.38	they	187.77	please	120.05	good	109.90
rescuers	495.52	them	279.40	team	185.83	humanity	117.31	their	104.04

Table 2 rarely has the function words. The content word with the highest keyness was "boys" which was similar to the result form the word frequency list. A comparison of the results in Table 1 with Table 2, suggested that analysing words using the keyword analysis method gave a better result. However, the frequency should not be ignored, so the researchers removed the function words from the lists and combined the remainder, as discussed in the next section.

3. Lexical Profiling

In this step, the word frequency list and the keyword list were lexically profiled using AntWordProfile. The results are shown in Tables 3 and 4.

Table 3: Fifty most frequent words after function word deletion

Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq
boys	172	world	62	heroes	50	people	37	let	27
god	148	kids	59	just	46	please	34	save	27
rescue	108	bless	56	now	45	life	33	very	27
so	101	hope	54	thai	45	story	32	diver	26
news	93	praying	54	well	42	brave	31	help	26
prayers	77	great	53	hero	41	rescued	31	thailand	26
cave	76	children	52	cnn	40	sad	30	lives	25
good	68	safe	51	rescuers	39	love	29	rest	25
coach	64	team	51	best	38	pray	29	rip	25
thank	62	divers	50	involved	38	know	28	see	25

Table 4: First fifty highest log-likelihood (LL) value keywords after function word deletion

Word	LL	Word	LL	Word	LL	Word	LL	Word	LL
boys	1313.02	news	494.12	rip	232.94	amazing	144.18	children	111.32
rescue	999.91	coach	472.79	brave	227.90	safely	139.92	great	110.93
prayers	787.09	thai	467.27	thaicaverescue	226.37	oxygen	134.08	story	110.65
god	768.76	heroes	460.29	pray	220.61	let	121.89	good	109.90
cave	739.31	kids	405.80	thailand	217.02	please	120.05	scuba	103.66
bless	604.86	thank	308.46	hope	215.26	humanity	117.31	tanks	101.74
praying	579.14	hero	290.46	team	185.83	involved	117.18	twitter	96.89
divers	564.66	safe	288.38	sad	180.44	world	116.61	amen	95.06
cnn	545.42	rescued	276.18	everyone	147.73	save	115.40	selfless	94.45
rescuers	495.52	diver	272.23	awesome	145.71	caves	114.42	luck	93.30

After the function words had been deleted from the lists (leaving the first 50 content words with the greatest frequency and the highest keyness), the words contained in the two lists were combined and the results are shown in Table 5.

Table 5: Combination of the fifty most frequent words and the first 50 words with the highest keyness

No.	Word	No.	Word	No.	Word	No.	Word	No.	Word
1	bless	14	heroes	27	rip	40	Best	53	people
2	boys	15	hope	28	sad	41	Caves	54	rest
3	brave	16	involved	29	safe	42	everyone	55	safely
4	cave	17	kids	30	save	43	Help	56	scuba
5	children	18	let	31	story	44	humanity	57	see
6	cnn	19	news	32	team	45	Just	58	selfless
7	coach	20	please	33	thai	46	Know	59	so
8	diver	21	pray	34	thailand	47	Life	60	tanks
9	divers	22	prayers	35	thank	48	Lives	61	thaicaverescue
10	god	23	praying	36	world	49	Love	62	twitter
11	good	24	rescue	37	amazing	50	Luck	63	very
12	great	25	rescued	38	amen	51	Now	64	well
13	hero	26	rescuers	39	awesome	52	oxygen		

The words in Table 5 are the result of combining the results of the word frequency list and keyword list producing 64 words. This showed that more than half of the words occurred in both the word frequency and keyword lists. Since these 64 words were mostly used and were the keywords in the comments, the researchers analysed and annotated them using LIWC and Semantic Tagger, respectively, to determine whether the commenters' attitudes were positive or negative.

4. Dispersion

Before discussing the attitudes of the commenters, the 10 most frequent words and the first 10 words with the highest keyness were compared and combined (Table 6)

Table 6: Comparison and combination of the ten most frequent words and the first 10 words with the highest keyness

First ten most frequent words		First 10 keywords		Combination
boys	prayers	bless	cnn	boys, god, rescue, so,
god	cave	boys	coach	news, prayers, cave,
rescue	good	brave	diver	good, coach, thank,
so	coach	cave	divers	bless, brave, children,
news	thank	children	god	CNN, diver, divers

After combining them, there were 16 words. However, when considering the meaning, “boys” could be grouped with “children” and “diver” could be grouped with “divers”. Furthermore, CNN is the proper name of the news agency and it could be predicted that CNN would distribute only data collected from CNN, not BBC and so CNN was ignored here. See the dispersion plots in Figure 1.


Word
Dispersion plot
boys
children
god
rescue
so
news
prayers
cave

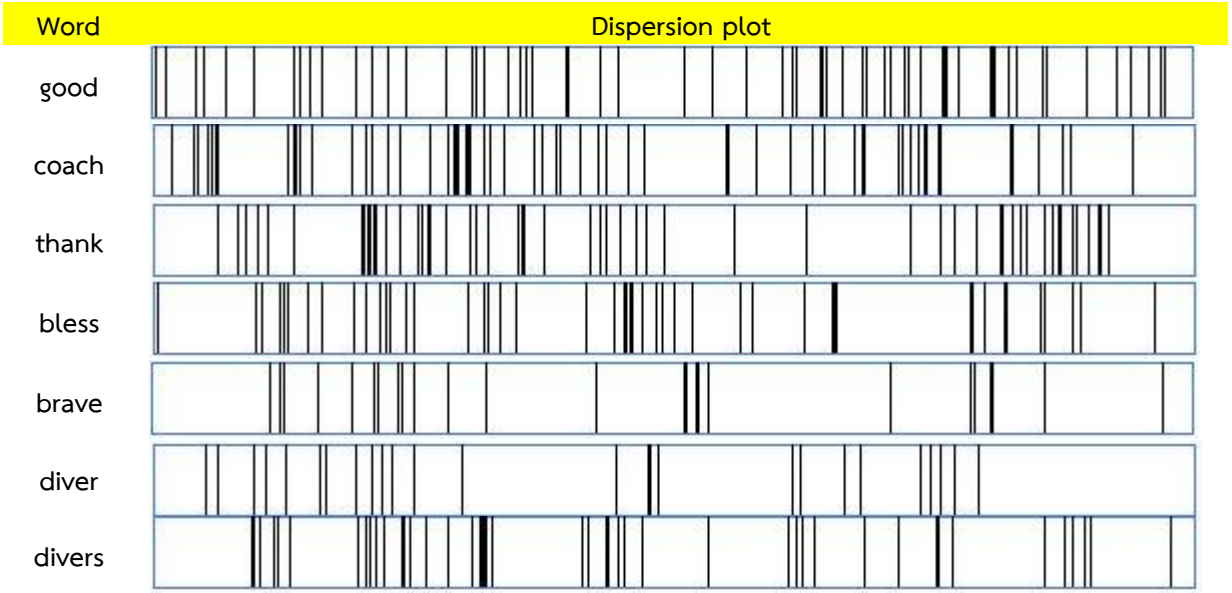

Figure 1: Dispersion plots of the 16 words across the corpus

From Figure 1, it could be concluded that “boys”, “god”, “rescue”, and “cave” were the topics mostly mentioned as it can be seen that the number of the lines appeared in bigger number, compared to other words in the plots. This reflected that people were concerned about the boys in the cave. The words “god” and “rescue” also reflected that they hope that the god would help the boys and relevant people like their families and rescuers. In addition, commenters talked about rescuing them. This was supported by the words “bless”, “prayers”, and “thank” that were often relevant or collocated with “god”. They might not occur as often as the previous set, but still, they appeared across the corpus (see the lines that disperse throughout theirs box). See some extracted comments below.

Thread 2: If a navy diver die on a conditions like that how the **boys** can make it if they had to dive...means the last option had been prove ... hope they find another way to get out there .... maybe a mine experts can help about some options.

Thread 3: Praying that all the boys & coach get out of the cave safely ????& all of the divers & rescuers ????**God** bless them all

Thread 4: We hope the best results on that **rescue** best wishes and keep strong our pray from Japan ???? #thailandRESCUE had just began #hope #recue #Thailand #Viral #makeitviral

Thread 5: Such a exciting story to follow, finally all well with the boys of **cave**.

Thread 6: We all know **cave** can be deadly dangerous But it fun to explore and out of the heat

Thread 7: God **bless** those kids

Thread 8: God **bless** them

Thread 9: **Prayers** to these boys and their families.

Thread 10: **Thank God**! Hopefully they all get out safe.

5. Attitudes towards the Thai Cave Boys Incident

To analyse the attitudes of the commenters towards the incident of the Thai cave boys, the combination of the most frequent words and the words with the highest keyness, (64 words in total) were analysed and annotated using LIWC and Semantic Tagger, respectively. The results from the two programs were then compared (Table 7).

Table 7: Overall analysis results using LIWC

Traditional LIWC Dimension	Score
Social words	18.81
Positive emotions	28.10
Negative Emotions	1.60
Neutral Emotions	70.30

From Table 7, among the positive, negative, and neutral emotions, the words belonging to neutral emotion were used the most (70.30 per cent). The proportions of positive and negative keywords used in the comments were clearly different (28.01 per cent and 1.60 per cent, respectively). This indicated that, from comparing the positive with the negative emotions, commenters had more positive attitudes towards the incident. Some of the words expressed to show the positive emotions towards the incident were “love” and “good”. Meanwhile, some negative emotion words were “fuck” and “shit”. Since most of them harmonised and commented with positive attitudes, they communicated through comments and replied with each other building a virtual community. And to do that, they used social words like “friend, family and us” to build a relationship with other users. See some extracted comments below.

Thread 11: *Love* the wild boars team

Thread 12: Welcome to be Thai citizen and hope this action to give *good* opportunities to these children.

Thread 13: Has anyone asked the coach yet "What the *Fuck* were you thinking?"

Thread 14: Watching that shit they probably wished they were back in the cave.

Thread 15: I hope after all of the dust settles and the boys and their teacher are able to recover from their ordeal they will do something kind and generous for the *family* and *friends* of the diver who died trying to save them.

Thread 16: On behalf of the Thai people I would like to say BIG THANK to all people for your help & hand that you give *us*.

Table 8: Overall results of annotation using Semantic Tagger

Value	Score
Positive words	56.25
Negative words	3.12
Neutral words	40.63

Since the Semantic Tagger is an annotation program, it ignores the context and focuses only the possible meaning of each word. Rather than relying on only the program itself, the researchers asked three linguists for agreement to confirm the results of the semantic tagger. When they had any disagreements on annotation, the linguists were again asked to negotiate the final annotation results. The final results (Table 8) showed that 56.25 per cent of the keywords were positive words, followed by 40.63 per cent as neutral words. Only 3.12 per cent of the keywords were negative words. This meant that people commented in a positive way towards the Thai cave boys incident. Some extracted comments with the Semantic Tagger results are provided below.

Thread 17: I pray that they all be rescued [S8+].

Thread 18: To the rescuers [S8+], best [A5.1+++] of luck [X9.2+]. From Mexico, prayers [A7+] will continue to go your way!

Thread 19: I've been following this since day 1 and just praying [A7+], so keep bringing them out. Everyone is so wonderful that's the way it should be everyone helping [S8+] each other. Love [E2+] and caring its so wonderful.

Thread 20: Continuing praying [A7+], I have 3 grandchildren can't imaging the agony!

Thread 21: So heart-warming to see people working together respectfully to help [S8+] others.

Thread 22: Million Thanks [S1.2.4+] to all of you and everyone who's been helping [S8+].

As you can see that [+S8] appears in Thread 17, 18, 19, 21 and 22. [S] refers to any words categorised in the semantic field involves SOCIAL, ACTIONS, STATES AND PROCESSES. [8] refers to the sub-semantic field of [S] involves helping/hindering. [+/-] refers to helping (+) or hindering (-). Since "rescue" in Thread 17 and 18, and "help" in Thread 19, 21 and 22 are related to helping, they were annotated as [+S8]. (For more detail about semantic fields used in the Semantic Tagger program, see http://ucrel.lancs.ac.uk/usas/.)

The results of the two programs were correlated. Commenters were worried about the lives of the boys and the coach. They prayed, hoped, and supported each other in the best way they could, to cheering themselves up and praying to the gods to support the rescuers and the boys in the case of commenters. Also, people worldwide viewed this incident as a good chance to build and improve on relationships between people from different nations. See the following extract.

Thread 23: This is the first time the whole world is supporting the same sports team and it is just beautiful.

Discussion

The frequency and keyness results indicated that words like “boys”, “rescue”, and “cave” were ranked in the top 10 words, indicating that the commenters were not out of the topic. What they mainly talked about and remained focused on was still the incident. The word “boys” was the most mentioned sub-topic suggesting that people were worried about the boys and frequently talked about them. The sub-topics “rescue” and “cave” were next most popular, respectively. Commenters were concerned about the process and the result of rescuing the boys. Furthermore, they, many times, discussed the cave in many issues, for example general information about the cave, knowledge about the cave, and experience of being in a cave. When comparing the results with other studies on the language used in online media such as Drasovean and Tagg (2015), Tsou et al. (2014), and Ivković (2013), the comments in the current study mostly stuck to the topics which were similar to comments on TED.com (Drasovean & Tagg, 2015). This implied that the Twitter users mostly focused on the topics, agreed with other users and were rarely out of the topic. On the other hand, Tsou et al. (2014), and Ivković (2013) studied comments on Youtube.com. Their results supported each other that people mostly commented interactively, was less focused on the topic, and there was often a lot of conflicts among users. Regarding the current topic the results indicated both similarities and differences. Similar to the Youtube studies, the commenters of the Thai cave boys incident news interacted with other users using social words because most of them were worried about the boys and wanted to support spiritual hope. As a result, they agreed to pray for the boys. The results also reflected that people felt so happy that people from all over

the world were trying so hard to support each other in this incident. As they could not go and help the boys themselves, they chose to support by offering words to cheer up those involved and by praying for all relevant people.

In contrast to the Youtube studies, they stuck to the topic (the Thai cave boys) and did not deviate from the main theme like the Youtube's commenters did. Furthermore, the results in the current study correlated with Herring (2008) where she summarized three main characters of online communities as: the online users share a common interest; there is a continuous existence of interactions; and the usage of emotional tones in communicating with others. In the case of the Thai cave boys incident, Twitter users were commonly interested in the survival of the boys and they interacted with each other using social words and an emotional tone.

Apart from hoping and praying for the wellbeing of the boys, there were some comments addressing other children's problem issues in other parts of the world. See some examples below.

Thread 24: While the rescue going on the gov of Cameroun dictator biya and @ManuelMacron are women and children in a new Rwanda style genocide, 15 not 12 football players were killed in buea a city in Ambazonia, #GenocideInSouthernCameroons the @UN is silent

Thread 25: Humanity is the largest lie history has ever known. The children of Syria die every day from bombing and hunger and the world did not care for them while the children were trapped in a hut that the whole world had done for them. There is no justice

Thread 26: It seems like the Thailand government is working harder to reunite these children with their parents then the American government is trying to reunite separated immigrant children with theirs. Maybe because one sees them as children while another sees them as dirty invaders?

As you can see from the threads, some commenters raised the problems of the children in other parts of the world hoping that those children should have been helped like in the Thai cave boys case. The results from the attitude section showed that most of commenters had a positive attitude towards the incident since a large proportion of the language chosen involved positive words. The results from both LIWC and Semantic Tagger supported each other in this issue which strengthened the idea that the Twitter users had a positive attitude towards the incident. In contrast, negative words were rarely seen in the comments.

The relationship between the topics mostly mentioned in the comments and the attitude towards the incident found in the comments were in harmony. The most mentioned topic was worry about the boys and a strong hope they would be rescued safely. In addition, there were comments of happiness that people from different sides of the world shared the same desire—a wish the boys end up safe. Coincidently, most of the commenters had a positive attitude towards the worldwide cooperation and strongly hoped the boys would come out safely.

Sentiment Analysis is another instrument to measure people's attitude in addition to LIWC and Semantic Tagger. Sentiment Analysis has recently been of interest and applied as an evaluation method in research, especially concerning people's attitude and emotion, i.e. customers' attitude. Sentiment Analysis has a wide range of emotion evaluation from the two most basic emotions (positive and negative), to multi-emotions subdivided into fine-grained emotions (Pang, Lee & Vaithyanathan, 2002; Pang & Lee, 2004; 2005; Koppel & Schler, 2006). The researchers recommend future research on the use of Sentiment Analysis to evaluate attitudes and investigate more in-depth results.

The limitation of this study was that data were only collected from the BBC and CNN official accounts on Twitter, and since news agencies normally

establish news in the interest of people over a short period, the data were restricted to 16,621 tokens for the whole corpus. Collecting data from other online channels like Facebook or Instagram might lead to a broader view of people towards the incident.

In conclusion, considering the comments analysed, people worldwide had a good attitude towards this incident, hoping that everything would end happily with no sign of any aggravation regarding the incident. People around the world are still ready to lend a hand when the world needs it.

References

Allport, G. (1935). Attitudes. In C. Murchison (Ed.), *A Handbook of Social Psychology* (Vol. 2, pp. 798–844). Worcester, MA: Clark University Press.

Anthony, L. (2014). *AntWordProfiler Version 1.4.1* (Software). Tokyo, Japan: Waseda University.

Anthony, L. (2018). *AntConc Version 3.5.7* (Software). Tokyo, Japan: Waseda University.

Antonak, R., & Livneh, H. (1995). Direct and Indirect Methods to Measure Attitudes toward Persons with Disabilities, with an Exegesis of the Error-Choice Test Method. *Rehabilitation Psychology*, *40*(1). 3-24.

Baker, C. (2006). Introduction to Language Policy: Theory and Method. In T. Ricento (Ed.), *An Introduction to Language Policy: Theory and Method*. (pp. 210–228). Oxford, UK: Blackwell.

Clement, J. (2019). *Twitter: Number of Monthly Active Users 2010-2019.* Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*(2). 213-238.

Drasovean, A., & Tagg, C. (2015). Evaluative Language and its Solidarity-Building Role on TED.com: An Appraisal and Corpus Analysis. *Language@Internet, 12*(1).

Eagly, A., & Chaiken, S. (1993). *The Psychology of Attitudes.* New York: Harcourt Brace Jovanovich.

Gardner, R. (2002). Social Psychological Perspectives on Second Language Acquisition. In R. Kaplan (Ed.), *The Oxford Handbook of Applied Linguistics* (pp. 160–169). New York: Oxford University Press.

Garrett, P. (2010). *Attitudes to Language*. Cambridge, UK: Cambridge University Press.

Graham, D. (2014). *Key-BNC* (Software). Retrieved from http://crs2.kmutt.ac.th/Key-BNC/

Gries, S. (2010). Useful Corpus-linguistics Statistics. In A. Sánchez & M. Almela (Eds.), *A Mosaic of Corpus Linguistics: Selected Approaches* (pp. 269–291). Frankfurt am Main: Peter Lang.

Herring, S. (2008). Virtual Community. In L. M. Given (Ed.), *Encyclopedia of Qualitative Research Methods* (pp. 920-921). London: Sage.

Herring, S., Stein, D., & Virtanen, T. (2013). *Pragmatics of Computer-Mediated Communication.* Berlin, Germany: De Gruyter Mouton.

Ivković, D. (2013). The Eurovision Song Contest on YouTube: A Corpus-based Analysis of Language Attitudes. *Language@Internet, 10*(1).

Kennedy, G. (1998). *An Introduction to Corpus Linguistics.* London, New York: Longman.

Koppel, M., & Schler, J. (2006). The Importance of Neutral Examples for Learning Sentiment. *Computational Intelligence,* 22(2), pp. 100–109.

Lindquist, H. (2009). *Corpus Linguistics and the Description of English.* Edinburgh: Edinburgh University Press.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 79–86.

Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the Association for Computational Linguistics (ACL)*. pp. 271–278.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the Association for Computational Linguistics (ACL).* pp. 115–124.

Pennebaker, J., Booth, R., & Francis, M. (2007). *Linguistic Inquiry and Word Count: LIWC* [Computer software]. Austin, TX: LIWC.net.

Rayson, P., Archer, D., Piao, S., & McEnery, T. (2004). The UCREL semantic analysis system. In *proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation* (LREC 2004), 25th May 2004, Lisbon, Portugal, pp. 7-12.

Ryan, E., Giles, H., & Hewstone, M. (1988). The Measurement of Language Attitudes. In U. Ammon, N. Dittmar, & K. Mattheier (Eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society* (pp. 1068–1082). Berlin: Walter de Gruyter.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Blackwell.

Tausczik, Y., & Pennebaker, W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology,* 29(1), 24-54.

Tsou, A., Thelwall, M. Mongeon, P., & Sugimoto, C. (2014). A Community of Curious Souls: An Analysis of Commenting Behavior on TED Talks Videos. *PloS ONE, 9*(4).

West, M. (1953). *A General Service List of English Words.* London: Longman.