

# A Corpus-Based Study Of The Representation Of Police, Protester, And Student In Thai Anti-Government Protest News

Akkarapong Worarattapong<sup>1</sup> , Supakorn Phoocharoensil<sup>2</sup>

<sup>1</sup>Language Institute, Thammasat University – Thailand, [akkarapong.wor@gmail.com](mailto:akkarapong.wor@gmail.com)

<sup>2</sup>Language Institute, Thammasat University – Thailand, [supakorn.p@litu.tu.ac.th](mailto:supakorn.p@litu.tu.ac.th)

## Abstract

The pro-democracy protest movement in Thailand in 2020 has generated considerable public interest. Media representation of protest movement contributes greatly to the perception of people towards any event or even persons. This study was conducted to explore how frequent social actors shared between two presses of Prachatai English Corpus (PCT) and Bangkok Post Corpus (BKP) were represented. All online news articles available were collected: BKP is composed of 363 news articles and PCT is of 136 news articles. The results of collocation analysis revealed that both presses described police, protester, and student differently. Police were regarded as violent actors in PCT, whereas BKP shows that police actions were not apparent. The protester was labelled as anti-government in BKP which is distinct to the portrayal by PCT which viewed a movement as for democracy. The prominent group of student protest was represented and personalized with its own title in PCT as bad student while in BKP bad student was used as an appositive. The article engages the discussion of Van Dijk's (1998) ideological square to demonstrate the bias existing in media representation.

**Keywords:** corpus-based, social actors, news, collocational analysis, discourse analysis.

## I. INTRODUCTION

In the year of 2020 saw an influential political movement led and joined mainly by students. This political gathering has sparked public interest as since 2014, coup d'état was staged to topple "the elected government of Yingluck Shinawatra, sister of former prime minister Thaksin Shinawatra. Its leader Prayuth was appointed by the coup-installed assembly to serve as the country's new Prime Minister" (Kongkirati & Kanchoochat, 2018). National Council for Peace and Order (NCPO) led by Prayuth seized the governmental power and was designated as a Prime Minister. Thailand has been ruled by military government since then. Government bodies and members of parliament were critically questioned of their designation. In 2019, the first election after the 5-year of military governance existed. The result of the election was that the prime minister Prayuth continued his tenure. The election was critically claimed by the public. However, the continuation of Prayuth was claimed to be unfair, mainly because of the nepotism as stated:

Prayuth appointed 36 of his former subordinates, seventeen of his classmates, and his own younger brother to positions in the NLA. Prawit Wongsuwan, another key NCPO leader and defence minister, appointed two of his younger brothers to the NLA. Besides Prayuth's and Prawit's families, four other leading military families are strongly represented in the NCPO and NLA: Ruddit, Kerdpol, Sitabutr, and Nakwanich (Kongkirati and Kanchoochat, 2018)

A range of factors which contribute to the emergence of pro-democracy protests can be identified. In terms of ideological concept, democracy and its principles were not fairly practiced and fully realized. It can be seen that often is the government regarded as dictatorship as opposed to democratic processes. The position of Prayuth as the prime minister along with its bodies is questionable due to the system of the appointed House of Senate. The demand for the new constitutional draft was called for by the protests. In terms of economy, it has been claimed by the protesters that the economy

become less successful because the inefficient administration of the government. The global pandemic of COVID-19 has catalyzed the worsened part of Thailand's economy which was believed to highlight the unsuccess of administration and management. The key aspects, to the researcher's belief, are the societal factors. Various forms of oppression upon the people were realized. The direct oppression can be seen from the abduction of people who expressed different political views, the charges filed against the number of activists, and the confinement of the leaders of the protests. Towards indirect oppression, the movement calls for social equality in a wide range of social aspects, from gender equality to students' rights. The protester's stances and viewpoints are presented with three commands (1) dissolution of parliament, (2) a stop to the intimidation of people, and (3) the drafting of a new constitution. ("Three Demands of Thai Protest Movement," 2020) These are the earlier proposed commands. Consequently, three more immediate commands were (1) the resignation of Prayuth Chan-o-cha, (2) the drafting of a new constitution from the people (3) the reform of monarchy.

The role of media especially presses or news is undeniably important as it is said that media coverage is able to contribute to the gaining of traction, earning public support and even influencing governmental authorities (Gamson and Wolfsfeld, 1993). News is argued as it is not a simple presentation of factual information. It is rather a contested area for ideological stances or a site of meaning construction, as stated "news is a representation in this sense of construction; it is not a value-free reflection of 'facts.'" (Fowler, 1991, p.4) News can serve as a political tool to produce or reproduce certain ideas or a set of beliefs. A number of studies have been conducted to explore news discourses in various aspects. Discourse analysis facilitates the study of at the structure, the rhetoric, or styles of news articles, while some researchers can focus on the close inspection in texts to seek how the text reinforces or undermines ideology. Recent approach to discourse analysis involves corpus linguistics into the study, which brings fresh perspectives on news study as it illustrates cumulative effects of media power. To the scope of this article, adopting a corpus-assisted discourse analysis, the research places the focus upon the representations of key social actors—police, protester, and student because their roles in strikes or confrontation are prominent in any protests. The research questions have been

developed to identify the representation, similarities, and differences of the social actors police, protester, and student in 2020 Thai anti-government protest news.

1. How are the social actors police, protester, and student represented in the anti-government protest news written by Prachatai English and Bangkok Post?

2. What are the similarities or differences of how the social actors police, protester, and student are represented in the anti-government protest news written by Prachatai English and Bangkok Post?

## II. LITERATURE REVIEWS

### NEWS DISCOURSE

As the present study aims to explore the news of Thailand's anti-government protests occurring in the year of 2020. Events of the protests have been extensively reported in news especially in online platform. The protest movement in Thailand saw a tie to media coverage as it serves as a mediated channel to represent the social events and even social actors. The role of media discourse or news discourse can be regarded as a key influence in society as stated by Bednarek and Caple (2012) as follows:

Sometimes we may modify our behaviour, ideas and beliefs, based on what we have read or heard and in this way we navigate our paths through our lives as members of a particular social group. This also points to the power of the media: the influence they exert both on our governments and major institutions as well as their ability to shape our ideas and behaviors. (p .6)

It can be seen that news discourse or language in the news is critically important to the study of language as it greatly affects almost all members in a particular society. The influence of media or beliefs suggest that ideology or a set of beliefs can be mediated through news discourse. Media can be regarded as ideological brokers producing and reproducing discourses of dominant social order (Fairclough (1995) and Van Dijk (1988) cited in Brindle, 2015).

Another important aspect in news discourse is that it carries constructive elements. The media can construct people, events, or places to serve a particular ideology. It is useful to embark on the definition of discourse and discourse analysis which are barely uniformly defined. Discourse analysis emerged as a perceived debunk to the overemphasis on grammatical study on a language within a single sentence. Harris (1952), having observed how the meaning can be

conveyed beyond a single sentence to arrive at the accurate understanding of a text, maintained “Beyond proximate sentences or textual episodes may be necessary, where themes and ideas are fully developed and brought to closure.” (p. 474). Hence, discourse can be said to a move away from a focus on meaning in words and sentences to a broader level of meaning beyond a sentence.

Discourse engages with the representation as stated by Fairclough (2015) as “ways of representing aspects of the world – the processes, relations and structures of the material world, the ‘mental world’ of thoughts, feelings, beliefs and so forth, and the social world.” (p.124). Moreover, discourse can be constructed as it does “not only represent the world as it is (or rather is seen to be), they are also projective, imaginaries, representing possible worlds which are different from the actual world, and tied in to projects to change the world in particular directions.” (p.124). Therefore, according to the Fairclough’s definition, the study of discourse plays a great role in uncovering the construction of language. News then serves as a great arena for seeking how events, actions, persons, or places are described differently by a variety of different news presses.

### **SOCIAL ACTORS IN NEWS**

In news discourse, to adopt the viewpoint that news can construct a variety of elements in presentation. Social actors can be subject to the analysis of news discourse. This presents study adopts the conceptualization of social actors posited by Theo van Leeuwen. He has interwoven different types of social actors discussed in the arena of critical discourse analysis (CDA). In Leeuwen’s (2008) prominent chapter of “The representation of social actors” in *Texts and Practices*, he began the introduction posing the focal questions of “what are the ways in which social actors can be represented in English discourse? Which choices does the English language give us for referring to people?”. In particular, the work exposes the readers to the racist discourse representing immigration in association with fear, involving “the fear of loss of livelihood and the fear of loss of cultural identity as a result of the ‘influx’ of immigrants who are perceived as ‘other’, ‘different’ and ‘threatening’”.

Leeuwen (2008) constructs the elaboration of social actors by placing an emphasis on “sociological categories” rather than on

“linguistic categories”. In closing his writing, he affirmed that system network was formed covering “a variety of linguistic and rhetorical phenomena.” The reasons for focalizing on the “sociological categories” were provided. The lack of biuniqueness of language or one-to-one correspondence of a language accounts for the first explanation. In other words, sociological concept cannot be fully realized by a single linguistic phenomenon. Agency was exemplified as it can appear in more than one realization—possessive pronouns or prepositional phrase. The instances were provided as in ‘our intake of migrants’ and “People of Asian descent say they received a sudden cold-shoulder from neighbors and co-workers.” As can be seen, two forms of realizing the agency were presented. Leeuwen (2008) then aimed at uncovering the representation of social actors considering a set of relevant categories. Second, the categories should be regarded as a pan-semiotic covering a variety of ways meanings can be represented, which was described as followed:

A given culture (or a given context within a culture) has not only its own, specific array of ways of representing the social world, but also its own specific ways of mapping the different semiotics on to this array, of prescribing, with greater or lesser strictness, what can be realised verbally as well as visually, what only verbally, what only visually, and so on. (p. 34)

In discussing the categories, a large and diverse corpus of texts including fictional narratives, comic strips, news stories, etc. was introduced to be used as an example. It’s labeled as ‘Our race odyssey’ published in the *Sydney Morning Herald*, a broadsheet newspaper. The present research aims to discuss the following concepts: exclusion (suppression and backgrounding), role allocation (activation and passivation), genericization and specification, assimilation (aggregation and collectivism), association and dissociation, nominalization and categorization, and functionalization and identification (classification, relational identification and physical identification).

Social actors can be excluded and included in representations as “to suit their interests and purposes in relation to the readers for whom they are intended.” (Leeuwen, 2008, p.38) The exclusion can serve many functions ranging from the details which can be referred to or the intended strategy of propagandizing an idea.

Two versions of exclusions were offered: suppression and backgrounding. Suppression concerns the radical deletion of social actors while backgrounding is less radically realized when social actors are not mentioned as the direct carrier of a particular action.

Role allocation assigns the role of who is represented as an actor or the receiver of that particular action, as indicated 'agent' or 'patient' respectively. Representations can be argued to entail activation and passivation. Activation is associated with forces which are dynamic in an activity whereas passivation is realized when social actors undergo the activity or receive the end of it.

In genericization and specification, social actors can be generalized and referred specifically. Genericization can be realized with plural without article or singular with a definite or indefinite article. Mass nouns when used as a reference to a group in general are used without article. However, the use of tense can also determine the interpretation of generic reference in that habitual present tense was associated with general noun. Specific reference was presented in conjunction with numerative.

Social actors can be represented with assimilation that occurs when an individual is represented as in group which can be subcategorized into aggregation and collectivization. Aggregation refers to the way participants are treated as statistics. It can express the function of the majority to justify certain action or control social practices based on democratic principle of representing the majority of people's voice in association with facts in form of polls, survey, or research. Collectivization occurs in the use of the pronoun as in 'we' and collective term of community. For example, experts were collectivized as in the committee or surveys to show expression of agreement towards certain extent. They were mostly referred as an individual in which institutional affiliations were displayed.

In nominalization and categorization, social actors can be represented with nomination, highlighting their 'unique identity' whereas categorization deals with the representation of social actors in terms of "identities and functions they share with others." The realization of nominalization can be done by proper nouns that can be formal, semi-formal, or informal as could be only a surname marked with honorifics (Lieutenant Franklin), a given name and surname (Calvin Harris), and a given name (John).

Functionalization primarily describes the activity social actors perform, covering their occupation or role. Identification involves what they permanently are. It can be further divided into three subcategories: classification, relational identification, and physical identification.

Classification is processed by a given society or institution to allow differentiation of classes of people, including age, gender, race, or social class. The categories of the classification are not permanently fixed. They culturally and historically vary.

Relational identification represents social actors with regards to "personal, kinship or work relation to each other." The examples include 'mates', 'co-workers' or 'uncles'. The possessive forms can appear in a possessive pronoun, a genitive, or postmodifying prepositional phrase as exemplified by 'her mother', 'the child's father', and 'a father of four' respectively. Leeuwen (2008) further argued that the role relational identifications can be negative. To illustrate, the involvement of such a relationship as family relation can allow 'nepotism' or 'corruption.'

Physical identification describes social actors targeting at physical attributes which identify them in a context. The realizations could be in the forms of noun, adjectives, prepositional phrases, or generalized classifications as in 'blonde', 'short', 'with disability' or 'man' respectively. The critical viewpoint regarding physical identification shows that it cannot be distinctly distinguished between physical identification and classification. For example, the reference to the color of skin can be categorized in both physical identification and classification.

It is to be noted that to discuss a discreet list of all representations of social actors can be challenging; however, the aforementioned representations will primarily be adopted in analyzing the data. The context and the concept of the research study were provided. It is also important to discuss the key methodological approach to the examination of the representations of frequent social actors.

## **A CORPUS-BASED APPROACH TO DISCOURSE ANALYSIS**

A corpus-based approach to discourse studies is barely new in the field of language study. In general sense, it is often regarded as a marriage between two distinct methodologies: qualitative paradigm and quantitative paradigm. Corpus

linguistics is closely associated with numerical data dealing with a large collection of data to seek representative sample of a language while discourse analysis extensively zooms the focal lens into the text to explore the meaning conveyed beyond sentences. The observed difference was remarked by leech (2000, p. 678-680, as cited in mcenery et al., 2006) as followed:

while DA emphasizes the integrity of the text, corpus linguistics tends to use representative samples; while DA is primarily qualitative, corpus linguistics is essentially quantitative; while DA focuses on the contents expressed by language, corpus linguistics is interested in language per se (p.111)

The present study's methodology is influenced mainly by the prominent work by Paul Baker's "Using Corpora in Discourse Analysis" (2006). The section will discuss discourse analysis followed by the role of corpus linguistics in discourse analysis.

Discourse analysis emerged as a perceived debunk to the overemphasis on grammatical study on a language within a single sentence. It can be regarded as the method of studying patterns in text either in oral or spoken to form coherence and meaning, describing the operation of a language. It can deal with the communicative purpose of a language. The studies adopting this type of discourse analysis are predominantly genre or move analysis in which the structure and organization of the texts are studied to discover the relationship between text structure along with its functions (Gee, 2015).

Discourse analysis can also engage with the representation as stated by Fairclough (2015, p. 45) as "ways of representing aspects of the world – the processes, relations and structures of the material world, the 'mental world' of thoughts, feelings, beliefs and so forth, and the social world." Moreover, discourse can be constructed as it does "not only represent the world as it is (or rather is seen to be), they are also projective, imaginaries, representing possible worlds which are different from the actual world, and tied in to projects to change the world in particular directions." Therefore, according to the Fairclough's definition, the study of discourse plays a great role in uncovering the construction of language.

Corpus linguistics has been defined by numerous scholars, referring to "the study of language

based on examples of real-life language use." (McEnery & Wilson, 1996, p.1). A corpus can also be defined as "a collection of used language." (Brazil, 1995: p.24) The reference to 'used language' deals with the authenticity or naturally occurring language. Gearing towards a more modern and practical takes on corpus linguistics, McEnery and Hardie (2014) presented an involvement of a digital tool or device, defining it as followed:

We could reasonably define corpus linguistics as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions. The set of texts or corpus dealt with is usually of a size which defies analysis by hand and eye alone within any reasonable timeframe. It is the large scale of the data used that explains the use of machine-readable text. (p.1)

It is worth noting that corpus linguistics at the present is deemed inseparable of electronic device as the input of data is to be "machine-readable text." The advantages are also stated as the procedure of analyzing the data can be less time-consuming. In addition to its convenience, corpus linguistics is positively perceived as it can provide generalizability of research study.

Corpus data and the study of news discourse are complimentary as news is the arena where discourses are produced and reproduced, as suggested by Fairclough (1998, p. 45) that "the effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader and so forth." In other words, corpus linguistics allows the hidden message to be uncovered.

In conclusion, the present research aims to apply a corpus-based approach to study representation of social actors in anti-government protest news.

## RELEVANT STUDIES

Corpus linguistics can be utilized in the research study of protests in South East Asia which are Occupy Central or Umbrella Movement in Hong Kong and Sunflower Movement in Taiwan. The situation in Hong Kong is, in some ways, similar to that in Thailand where the movement is the political protest, Occupy Central in 2014. May L-Y Wong combined both corpus and cognitive approaches to investigate aggression of social actors in political protests represented in media. Two presses, China Daily and the South China

Morning Post (SMCP), perceived to hold opposing viewpoints in the political conflict are compared. To collect the data, Wong retrieved news articles from the online database ProQuest, search words pertinent the movement: Occupy Central and Umbrella Movement. Then, 437 articles from China Daily and 438 from SMCP were analyzed by means of the concordance program, AntCont. Initially, collocation analysis was performed with police as the search word with the span of 5L-5R. Thereafter, linguistic instantiations in corpus examples were examined to explore discursive strategies and ideology. Framework of analysis was based a cognitive-linguistic focusing on media representation, in which “discourse always serves as a prompt for an array of conceptual processes.” (Wong, 2017). Four linguistic strategies, proposed by Hart (2015), are employed in analysis: structural configuration, framing, identification and positioning. They deal with the representation of basic event structure (domains of experience such as action, space, force, and motion), the categorization of an entity leading to patterns of inference and evaluation (frames such as success, victim, or failure), the presented and selected aspects of the scene facilitated by nominalization resulting in loss of a certain amount of information. Yang engaged ideological square to discuss the findings in the conclusion. China Daily and SMCP display, towards the same event of Occupy Central, different picture: China Daily reports positive presentation of police whereas SCMP shapes the image of police negatively. To give an illustrative example, it is evident in the analysis of framing or the categorization of identity. Through success frame, police were pictured with positive words as ‘applauded’ as in (1) by China Daily, while in (2), SCMP used the verb phrase of ‘widely criticized’ through failure frame to engage the negative image of police.

(1) Police officers assisting the clearance operation in Admiralty deserve to be applauded for their professionalism and common sense (China Daily, 12 December 2014).

(2) Police have been widely criticized for their handling of a protest largely portrayed as peaceful (SCMP, 26 October 2014).

In Taiwan, Brindle (2015) stresses the vital role of media and news discourse in reproducing the discourses of dominant social rules. He further argued that patterns of particular language usage contribute to interpretational patterns for the understanding of social issues, relations and identities. The scope of his study deals with the

Sunflower Student Movement in Taiwan, focusing on how the frequent words are constructed by discourse in news articles from two presses of The China Post and Taipei Times. Adopting the software of Sketch Engine, the analyst deals with the main features of corpus linguistics including the frequency list, collocation, and concordance lines to investigate the discursive constructions. Wordsmith Version 6 facilitates the keyness analysis to show the comparison of both corpora. The results show that The China Post negatively constructed the movement mainly as an impediment to the economy while Taipei Times presented more positive portrayal of the protests, stating that it serves as a struggle for democracy.

Brindle’s (2015) analysis was highlighted with the close inspection of the frequent non- function words generated from both corpora of the presses, The China Post and Taipei Times. Almost all words are similarly shared by both corpora. However, the word, Sunflower appeared only in the frequency list of Taipei Times corpus, in which he argued that The China Post did not prefer to associate the movement with the past democracy reform protest. It can be seen that the disappearance of the word proves to be significant in the study of discursive construction.

In addition, the subsequent analysis of each word from the frequency yields interesting results in relation discursive construction of the movement. Collocates and concordance lines were the driving tools of his analysis of the frequent words. To exemplify, Taiwan showed the highest frequency in both corpora, yet it yields differing representations of the movement. Through the analysis of Taiwan, The China Post represented the protests negatively. The object collocates of Taiwan includes hurt (13), visit (8), protect (6) and affect (6). For instance, hurt Taiwan’s economy is found as in “the prolonged demonstration is likely to hurt Taiwan’s economy, which only began to show improvement recently.” In the concordance line, the movement is portrayed negatively as it contributes to “a detrimental effect on the Taiwanese economy.” Besides, the grammatical feature of possessive form affirms the negative portrayal of the movement. The collocates are economy (22), future (11), democracy (11), development (7), history (5) and competitiveness (5).

The concordance “Taiwan’s economy ‘will weaken day after day’ until its collapse as its trade with China and the rest of the world

continues to shrivel because of the delay of the trade agreement going into effect.” (Brindle, 2015) shows the important factor to Taiwan’s economy which is the trade pact or trade agreement with China accompanied by the blame on the protesters as causing the delay “by the legislature occupation.” Therefore, with the analysis of collocation and grammatical feature, The China Post negatively constructs the image of the protest.

On the other hand, Taipei Times produced a positive image of the protest mainly as a struggle for democracy. The same word, Taiwan, was analyzed and the object collocates were safeguard (9), annex (6), protect (7), hurt (6), help (7), visit (6), defend (6) and support (6):

‘We owe these kids a better Taiwan’, said Chiu Shih-lung, who is in his 70s. Chiu, who came from Greater Kaohsiung, said he supported the students’ cause of safeguarding Taiwan’s democracy and demanding the withdrawal of the cross-strait service pact because ‘it endangers national security and sovereignty.’ (Taipei Times, 2014)

To illustrate safeguarding Taiwan’s democracy shed positive light on the protest. Unlike the China Post, Taipei Times depicts the pact as “a threat to democracy, national security and sovereignty.” Moreover, the analysis of possessive form of Taiwan supports the movement. The most frequent collocate in possessive form of Taiwan is democracy (47) as in the concordance line: “our parents have failed to safeguard Taiwan’s democracy for us, but we would rather die than let our children grow up in a land without democracy.” It can be seen that the old generation are constructed as holding the responsibility to protect the democracy. The younger generation, on the other hand, is portrayed as “mature, wise and committed to a transparent democratic society.” (Brindle, 2015, p. 12).

Keyness is also one of the interesting features of the study. When compared against each other, the labelling strategy was discovered. Both corpora refer to the political party differently. The China Post refers to the party as Kuomintang as to construct the party that “accepts a ‘One China Principle’.” Taipei time, however, label the party as Chinese Nationalist Party to highlight the fact that the party derives “from the mainland China that imposed itself on the Taiwanese people when the defeated nationalist

army retreated to Taiwan in 1949.” (Brindle, 2015, p. 14).

Reviewing the relevant studies, it can be seen that a corpus-based approach to study a variety of facets in media was extensive. However, in Thai’s context, little research adopting a corpus-based approach to the representation of social actors as a focal point has been conducted. It is also noteworthy that three key elements shared amongst previous studies are frequency list, collocates, and concordances which will be the main tools in analysis as well.

### III.METHODOLOGY

The combined methodologies of corpus linguistics and discourse analysis primarily serve as a key approach. This methodology carried varied realizations. It can be regarded as a corpus-assisted discourse analysis while some researchers refer to it as a corpus analysis of discourse. However, the present research will adopt the title of corpus-based as it “is used as an umbrella term for a range of corpus enquiries” (Flowerdew, 2012) The scope of the present research aims to study the representation of police and protester in the anti-government protest news as they ranked the highest in the normalized frequency or frequency per millions. In Prachatai English (PCT), the frequency per million of police is 6,604.40 with the raw frequency of 769. The frequency per million of protester is 4,438.23 with the raw frequency of 514. In Bangkok Post (BKP), the frequency per million is 7,516.43 with the raw frequency of 1,361. The frequency per million of protester accounts for 6,047 with the raw frequency of 1,095.

### DESCRIPTION OF THE CORPUS

The corpora which will be examined in the present research will include all news articles on the topic of anti-government protests. As the situation was recent at that time, all news articles are to be included. The corpora will be subdivided into two corpora: Prachatai English Corpus and Bangkok Post Corpus shortened as PCT and BKP. As both presses represent viewpoints from Thailand, all news articles about anti-government protests in British Broadcasting Company or BBC were collected in comparison as well.

Collected by the researcher, the news pieces were separated by folders and titled with date of the publication followed by the headlines of each news as exemplified by 20200920-Protest marchers plan 'big surprise'. The period of data

collection of all corpora started from October 2020 to March 2021. The size of the corpora are as follows: Prachatai English corpus (PCT) is composed of 136 news articles, counted as 115,812 tokens and 101,424 words. Bangkok Post corpus (BKP) is composed of 364 news articles, counted as 181,070 tokens and 160,382 words. British Broadcasting Corporation is composed of 36 news articles, counted as 27,499 tokens and 24,237 words.

### **CORPUS TOOL**

Sketch Engine, a web-based corpus tool, was utilized as the key instrument in this present study. Three major functions were used: wordlist, Word sketch, and concordances. Wordlist, or referred as frequency list, arrange words or lemmas (the basic form of words) according how frequent they appear in a corpus. The researcher drew a list of frequent social actors to be used as a search word for collocational analysis. Frequency can be represented as raw frequency or normalized frequency. In Sketch Engine, frequency or absolute frequency means “the number of occurrences or hits,” while normalized frequency means the hits as occurring per million which is useful in comparing corpora of different sizes. Normalized frequency is referred to as relative frequency or frequency per million. It is defined as follows:

A number of occurrences (hits) of an item per million tokens, also called i.p.m. (instances per million). It is used to compare frequencies between corpora of different sizes. number of hits: corpus size in millions of tokens = frequency per million The frequency per million is always related to the whole corpus or subcorpus, not to a text type. Restricting the query to one or more text types will affect the number of hits but the frequency per million will still be calculated using the number of tokens in the whole (sub)corpus.” (relative frequency, frequency per million)

To seek answer of how frequent social actors are represented, Word Sketch is a function that targets the examination of collocates or the words that tend to cooccur with the word. Collocation is important as it is “a way of understanding meanings and associations between words which are otherwise difficult to ascertain from a small-scale of analysis of a single text.” (Baker, 2006, p. 96).

Word Sketch provides collocational analysis separated by a range of categories, for instance, modifiers or X, verbs with X as subject, verbs with X as object, or nouns modified by X. In this study, the focus of analysis was upon the modifiers of a social actor as it can allow further exploration of how they are represented.

Concordances refer to “a list of all examples of the search word or phrase found in a corpus” (Concordance) Concordance analysis of the collocation was the final step to study in detail of how frequent social actors were represented. The image that the presses painted is detailed through concordances.

To provide an example, a wordlist was generated and police has the highest number of instances per millions which is 7,516.43 in BKP. Police was the search term in Sketch Engine. In BKP, police was mostly modified with riot. The collocation riot police showed that police was functionalized or represented with the role they are engaged in or what they are doing. Further concordance exploration of riot police was administered to see how they were described or represented. Data analysis included Leeuwen’s concepts of social actors and other discursive construction of police and protester.

### **FREQUENT SOCIAL ACTORS IN PRACHATAI ENGLISH AND BANGKOK POST**

Initially, drawing a list based on raw frequency is necessary “to give user a sociological profile of a given word or phrase enabling greater understanding of its use in particular contexts.” (Baker, 2006, p. 47) The function of wordlist in Sketch Engine facilitates the generation of list of frequent social actors in both corpora of Prachatai English and Bangkok Post. To determine the frequent social actors of two corpora with different corpus size, a normalized frequency or frequency per millions were used as a basis. In Sketch Engine, the frequency will be referred as instances per millions abbreviated as i.p.m.

Sketch Engine also allows the focus on lemmas as the words can be realized in many forms, as protesters and protester can be regarded as different words while lemma refers to “the basic form of a word, typically the form found in dictionaries” (Sketch Engine, n.d.).

**Table 1**

The Frequent Social Actors in the Corpus of Prachatai English



Lemmas	Frequency per million	Frequency
Police	6,604.40	769
Protester	4,438.23	514
Student	4,343.25	503
Activist	2,547.23	295
Officer	2,029.15	235
Anon	1,718.30	203
King	1,718.30	199
Parit	1,105.24	128
Criminal	1,027.53	119
Lawyer	906.64	105

**Table 2**

The Frequent Social Actors in the Corpus of Bangkok Post

Lemmas	Frequency per million	Frequency
Police	7,516.43	1,361
Protester	6,047.38	1,095
Leader	3,236.32	586
Student	3,153.48	571
Minister	2,595.68	470
Prayut	2,325.07	421
Demonstrator	1,977.14	358
Activist	1,535.32	278
Parit	1,308.89	237
Deputy	1,005.14	182

According to table 1 and table 2, police and protester are perceived as key actors in the protests because in any protest event skirmishes between police and protester often occur.

Deriving the frequent social actors police, protester, and student from the wordlist, modifiers of each search term will be further explored in order to seek the representation of social actors. In Sketch Engine, the function of collocations was referred to as Word Sketch.

Concordances of collocations were then closely examined. From this stage, the move from quantitative approach to a more qualitative approach was administered. Analysis and results will be described with the framework of Van Leeuwen (2008)'s theory of social actors.

#### IV. RESULTS

To answer the first research question of how the key social actors, police and protester, were

portrayed or represented in the protest movement, the collocates of both search words were extracted with Word Sketch in Sketch Engine. Extracted from both corpora Prachatai English and Bangkok Post shortened as PCT and BKP respectively, the collocates were presented in the tables. In each searched word, the modifiers of the search word were focused.

#### ANALYSIS OF THE SOCIAL ACTOR POLICE

Police stood at the highest frequent social actor in both corpora as of 6,604.4 (i.p.m.) and 7,516.43 (i.p.m.). Police is perceived to be a prominent social actor in any given protest movement as they are commonly associated with violent acts of confrontation and dispersal. It is found that the common collocates of police in any corpus are those associated with the role or the kind of police. In Prachatai English corpus (PCT), police is modified by control. Further exploration of concordances in which police appears shows that the full phrases are crowd control police and riot control police. In Bangkok Post (BKP), police is also referred to as riot police. The striking difference of how police is modified differently between the two corpora is that BKP tends to describe police with rankings while in PCT, police is described mostly as control police. Police was represented as the actor of violence in PCT.

In (1), the crowd control police was characterized as those carrying weapons as "gas masks and batons." PCT also used "apparently" to confirm the image, while in (3) the contrasting image was painted by BKP as the police was there "to ensure the security." In (2), PCT foregrounded riot police as the actor of beating the medical staff which portrays the riot police in a negative light. On the other hand, in (4), protesters were foregrounded as those throwing smoke bombs at the police. Sample concordances of social actor: police from PCT are shown in (1), and (2), while sample concordances from BKP are shown in (3), and (4).

- (1) At around 18.15, the protesters began marching towards German Embassy on Sathon Road, the entrance to which was blocked by crowd control police apparently equipped with gas masks and batons.
- (2) Meanwhile, the medical volunteer group Doctor and Nurse Association said one of their members was reportedly beaten by

riot police and detained, despite all members of the team wearing a vest identifying them as rescue worker.

- (3) Boarder patrol police and riot police were deployed to ensure security along the avenue.
- (4) Protesters then threw smoke bombs and bags of paint towards rows of riot police on the other side.

### **ANALYSIS OF THE SOCIAL ACTOR PROTESTER**

Protester is also the key social actor in the protest movement. Protesters were described and functionalized differently by PCT and BKP. In PCT, they were referred to as pro-democracy protester (327.28 i.p.m.) while they were mostly named as anti-government (706.91 i.p.m.) by BKP. Normalized frequency can confirm the difference in number between two corpora of different sizes.

Another important portrayal of protester as of innocent was present. Protester was described as peaceful and unarmed, probably to highlight the negative representation of the opposers. In (5), it expresses how the government is violating humans represented as peaceful. It is suggestive that the protester should not be meant for arrest and the government acts incorrectly. Furthermore, the government was implicitly criticized of imposing “excessive restrictions” and “treating unarmed protesters” with chemical-laced salvos. On the other hand, in BKP, that protesters were injured was rated as allegations or rumors as said in (7) and (8). It was noteworthy that the high-ranking police officers were defending the violent act of police. Sample concordances of social actor: protester from PCT are shown in (5), and (6), while sample concordances from BKP are shown in (7), and (8).

- (5) Thailand’s international friends should call on the government to stop arresting peaceful protesters, listen to their views, and allow them to freely and safely express their visions for the future.
- (6) However, the Thai government has “imposed excessive restrictions” against peaceful protest as well as treating unarmed protesters with chemical-laced salvos from water cannons, tear gas, and rubber bullets.
- (7) The deputy MPB commissioner also insisted that police did not use rubber bullets or live rounds on the protesters, as

alleged by some anti-government protesters.

- (8) Earlier in the day, Pol Maj Gen Yingyos Thepchamnong, spokesman for the Royal Thai Police, dismissed a rumor spreading among the anti-government protesters that the police were considering using bullets to kill them.

Earlier in the day, Pol Maj Gen Yingyos Thepchamnong, spokesman for the Royal Thai Police, dismissed a rumor spreading among the anti-government protesters that the police were considering using bullets to kill them.

### **ANALYSIS OF THE SOCIAL ACTOR STUDENT**

Appearing as a functionalized modifier of protester as student protester, the social actor student was listed as one of the most frequent actors in a wordlist. The normalized frequency 4,343.23 and 3,153.48 in PCT and BKP respectively. The protest movement in 2020 was different from the past protest as a number of the participants in 2020 were students. There were a number of flash mobs initiated by different universities in a series of days.

The frequent collocation is Bad student (224.5 i.p.m.) and (104.93 i.p.m.) in PCT and BKP respectively. The modifier bad does not refer to the adjective that means unpleasant. Bad student is the self-proclaimed title of an organization and a movement of students in calling for equality and fairness in education. The way PCT refers to the protesters for the betterment of education as Bad students clearly involves the use of personalization as the organization were given a human attribute. In addition, the title of this education protest movement was itself satirical as it is a mockery toward the viewpoint expressed by many supporters of government that the students active on the protest were sometimes as those who hate their own nations and bad. The student protesters, hence, harnessed the power of humor to question the moral quality of bad and good. In BKP, bad student was only used as referring to the organization in which the social actors will be pre-modified as “calling themselves” in (10). As from the samples, bad student was described differently. In (9), they were described as “best scripted” while in (10) they were eyed upon.

University student is another common collocation in both corpora. University student, despite informing the occupation of the protester as functionalization, serves as a description of a

person to express relational identification which represents social actors in terms of their work relation to each other, and it is realized by a closed set of nouns denoting such relations. As the protest movement was quite different in that the major participants were composed of university students such as Thammasat University, Chiang Mai University, and Kasetsart University as in (10). Showing the network of university group across the nation, the use of university student probably helps construct probably education identity to unite and relate both the audience and the protest goers belonging to the university itself and the network of university protesters.

In BKP, it is interesting that the title of Thammasat university was quite frequent in the representation of student as it is a proper noun of a Thai university renowned for political participation and activism as shown in the statement that “Thammasat University continued to play a key role in changes and development of democracy and politics in Thailand.” (Thammasat University) This was used as a nomination to describe the key social actors in the protest movements, for example, Mr. Parit and Mrs. Panusaya who are commonly known as the protest leaders. This echoes how BKP is inclined to nominate or detail key figures in the movement as discovered in how they nominated high-ranking police authority. In this case, Mr. Parit and Ms. Panusaya were recognized as the leaders of the protest as in (12). Using this nomination and details of education to describe Mr. Parit and Ms. Panusaya paints the image of leaders. To compare with PCT’s way of describing university students, PCT is more prone to provide space for university students in general and variably refer to different universities not just those associated with the key figure. Sample concordances of social actor: student from PCT are shown in (9), and (10), while sample concordances from BKP are shown in (11), and (12).

- (9) The protests by the Bad Students are thought to be the best-scripted and organized demonstrations in Thailand
- (10) Triam Udom and Kasetsart students rally against the government. Dozens of Triam Udom students and over 400 Kasetsart University students gathered at their schools to show their stance against the government
- (11) A source at the meeting said security agencies were to keep an eye on a group

of high school student activists themselves “Bad Students”, who have scheduled a rally for Saturday in front of the Education Ministry.

- (12) Mr. Parit and Ms. Panusaya, who are both Thammasat University students, and Mr. Arnon have been repeatedly denied bail by the court after their earlier releases saw them going back and recommitting the offences.

## V. DISCUSSION AND CONCLUSION

Analysis of the data from collocates from both corpora shows that the representation of shared frequent social actors differs as police were more likely to be detailed more in Bangkok Post and also described as injured where absent in Prachatai English. The image of police in Bangkok Post entails those of violent actors equipped with dispersal tools such as batons or gas. Most importantly, the act of police against protest event as in dispersal was justified and legitimized by the verb collocate of warn which serves as the key argument that the warning occurred before the start of any action.

Regarding the protesters, Prachatai English places the purpose of protest movements as a struggle for democracy as it commonly refers to the protesters as pro-democracy whereas Bangkok Post used anti-government as the frequent modifier along with the image of detainees or wrongdoers as evidenced by the frequent use of arrest and detain in BKP corpus, which is contrary to the representation in PCT in which the collocation arrested protester is written to explain and describe the cause for a protest in demand for freeing their fellow protesters.

Both presses of PCT and BKP, hence, serve as an arena for contested ideas and ideologies in which incongruence occurs. It can also be argued that both presses stand in different ideological stances. It is assumed that Prachatai English’s target audience are the supporters of pro-democracy movement. On the other hand, target readers of Bangkok Post are those who might express disapproving view of the movement. The explanation for differing representations can be discussed through the lens of ideological square. It revolves around the concept of the relationship between groups. The line separating ‘Us’ from ‘Them’ is the key characteristic of this concept. It is common that the others are more likely to be presented negatively as opposed to us who are better in terms in a variety of aspects. In general, the conceptual framework proposed

by Van Dijk (1998) in identifying ideology between the opposing groups is ideological square.

Four principles are (1) emphasize positive things about Us, (2) emphasize negative things about Them, (3) De-emphasize negative things about Us, and (4) De-emphasize positive things about Them. To cite examples, PCT constructs the image of police quite negatively through the reference of equipment while BKP allows the description of why police resorted to that action. It can be concluded that PCT put an emphasis on the negative elements towards the police while BKP de-emphasize negative things by justifying police actions.

The results of this study were quite in line with the previous studies as of Wong's (2017) work on the Occupy Central and Bridle's (2015) research on Sunflower Student Movement. It can be seen that there are some ideological discrepancies when representing social actors to suit the interest of news presses, showing support for either the government or the people. It is to be noted that the presses were influenced by outside force as the power of China which is unlike Thailand. This present study is primarily engaged with the presses from Thailand, yielding interesting results that within the same country, contrasting ideologies still persists. The stark contrast of viewpoints can be seen from the functionalization of protester as anti-government and pro-democracy. The struggle and yearn for democracy implied the persistent lack of it while the notion of defying government undermines the protest movement to the disapproval of the government.

In conclusion, the representation of the key social actors reinforces the idea that media or news are not the presentation of facts. It can be a tool to construct, shape, or paint the image of a certain group of people. A combined method of corpus linguistics and discourse analysis facilitate the exploration of insights into the representations of social actors catered by both prominent news presses of Prachatai English and Bangkok Post. The different perspectives could then be explained and discussed in the framework of ideological square proposed by Van Dijk (1998).

## REFERENCES

1. Baker, P. (2006). Using corpora in discourse analysis. Continuum.
2. Bednarek, M., & Caple, H. (2019). News discourse. Bloomsbury Academic.
3. Brindle, A. (2015). A corpus analysis of discursive constructions of the Sunflower Student Movement in the English-language Taiwanese press. *Discourse & Society*, 27(1), 3-19.
4. Fairclough, N. (2015). Language and power. Routledge.
5. Fowler, R. (1991). Language in the news: discourse and ideology in the British press. Routledge.
6. Gamson, W.A., & Wolfsfeld, G. (1993). 'Movements and media as interaction systems'. *Annals of the American Academy of Political and Social Science*, 528, 114-25.
7. Gee, J. P. (2005). An introduction to discourse analysis: theory and method. Routledge.
8. Harris, Z. S. (1952). Discourse Analysis: A Sample Text. *Language*, 28(4), 474-494. <https://doi.org/10.2307/409683>
9. Kongkirati, P., & Kanchoochat, V. (2018). The Prayuth Regime: Embedded Military and Hierarchical Capitalism in Thailand. doi:10.1017/trn.2018.4
10. Leech, G. (2000). Grammars of Spoken English: New Outcomes of Corpus-Oriented Research. *Language Learning*, 50(4), 675-724. <https://doi.org/10.1111/0023-8333.00143>
11. McEnery, A. and Wilson, A. (1996). Corpus Linguistics. Edinburgh University Press.
12. Sketch Engine. (n.d.). Collocation. Retrieved April 27, 2022 from [https://www.sketchengine.eu/my\\_keyword/s/collocation/](https://www.sketchengine.eu/my_keyword/s/collocation/)
13. Sketch Engine. (n.d.). Lemma. Retrieved April 27, 2022 from [https://www.sketchengine.eu/my\\_keyword/s/lemma/](https://www.sketchengine.eu/my_keyword/s/lemma/)
14. Sketch Engine. (n.d.). Glossary. Retrieved April 27, 2022 from <https://www.sketchengine.eu/guide/glossary/>
15. Taipei Times. (2014, March 31). Protests gather board support. <http://www.taipeitimes.com/News/front/archives/2014/03/31/2003586930>
16. Three Demands of Thai Protest Movement. (2020, December 17). Retrieved April 27, 2022 from <https://proximities.com/three-demands-of-thai-protest-movement>

17. Van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*. SAGE Pub.
18. Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Discourse Analysis*. Oxford University Press.
19. Wong, M. L. Y. (2017). Analysing aggression of social actors in political protests: Combining corpus and cognitive approaches to discourse analysis. *Journal of Aggression, Conflict and Peace Research*, 9(3), 178-194. doi:10.1108/JACPR-09-2016-0250