

CORPUS-BASED ANALYSIS: COLLOCATION OF NETIZENS' WORDS IN INDONESIAN NEWS IN 2020

Neng Sumiyati, Ely Fauziyah, Abdul Basid

Universitas Islam Negeri Maulana Malik Ibrahim Malang

Email: nengsumiyati537@gmail.com

Abstract: Netizens or internet citizens are defined as people who are active in online communities by being involved in several online forums or social media. In online media in Indonesia, netizens are often referred to as a group of the most righteous. This term is included in word collocation or word matching related to netizens. This study aims to describe the collocation of netizen words in Indonesian news in 2020. This research is a mixed method study. The primary data source of this study is news data contained in the corpus, namely Leipzig Corpora Collection. The results of this corpus-based research found that there are two types of collocations of netizen words in the Indonesian news in 2020. First, collocation of noun + adjective consists of the words curious, excited, very excited, envious, furious, amazed, excited, suspicious, concerned, and worried. Second, adjective + noun collocation consists of many, few, excited, angry, exciting, evil, negative, crowded, funny, furious, suddenly, excited, funny, surprised.

Keywords: *Corpus-based analysis, collocation, netizens' words, Indonesian news.*

INTRODUCTION

Today, the word media is blasting in this modern era. The media is a tool that makes it easier for someone to convey information. Many types of media are developing today, one of them is social media. Social media is online media which allow the users to share information, participate in discussion, communicate with others or post their opinions. Andreas stated that social media is a group of internet-based applications that build the ideological and technological foundations of web 2.0 and allow user-generated content creation (in Arlianti, 2020, pp. 77-78).

The use of social media has both positive and negative impacts. The positive impacts are easy to access the information, build friendships, create personal branding, promote products, and participate in social activities. While the negative impacts of using social media are being trapped into fake accounts, hostile behavior, or trapped into computer viruses (Rohmadi, 2016, pp.1-7). One of the other potentially damaging effects of increasing internet use is to reduce social relations and the quality of social interaction, particularly in local community organizations (Kraut, 1998, p. 53). In addition, there are other

problems associated with addiction. Internet users experience various changes or effects that occur each of their users when they use various online applications such as social networking sites, video games, cloud applications, and even online gambling (Cash, 2012, p 8).

As reported in tekno.kompas.com, Indonesia's population currently reaches 262 million people. More than 50 percent or around 143 million people have been connected to the internet throughout 2017. According to the latest report by the Indonesian Internet Service Providers Association (APJII), most internet users, as much as 72.41 percent, are still from the urban community. Its use has gone even further to communicate and buy goods, order transportation, do business, and work. The spread of the internet has risen to a new term in which we interact. In the real world, the people are known as citizens, and in the virtual world, the people are known as netizens (Sakti, 2020, p.33).

The word netizen refers to "citizen of the internet". This term is formed from two words: internet and citizen. In internet dictionaries, we find the notion of a netizen as a person or entity who is active in online communities or is active on the internet by being involved in several forums or social media. Thus, netizens are internet users who are active in communicating, expressing opinions, and collaborating on internet media (Romli, 2018, p. 180). According to Hauben (1996) "Netizen is not an internet user, but netizen is a word to describe anyone on the internet who supports the public and collaborates with the internet to help maintain and preserve the internet itself". In this case, Hauben gives the opinion that netizens are individual forms who become virtual world communities that have no virtual boundaries in which individuals can communicate without any limitations. In this case, the underlying reason is that netizens are considered active internet users in communicating and also expressing opinions and collaborating in any matter on social media (Goleman, 1999).

The word netizen in cyberspace in Indonesia is also often referred to as someone who is most correct for all his comments and opinions in addressing a problem or issue circulating in society. Therefore, researchers are interested in researching collocations or passwords that match the words netizens in Indonesian news in 2020. Researchers took data from Indonesian news in 2020 contained in the Leipzig Indonesian corpus because the data presented included big data. The Leipzig Corpus itself is the newest and largest corpus work from the Indonesian language which can be seen in the Corpora Collection of the University of Leipzig. This corpus, has more than 1.2 billion words. However, the Leipzig corpus data itself is only taken from the Indonesian language in online news media, websites, and Wikipedia (Kwary, 2015).

On the other hand, corpus or linguistic corpus-based analysis is very rarely studied. According to O'Keeffe and McCarthy Corpus Linguistics (Corpus Linguistic) is a method to analyze language patterns from large data sets empirically. Meanwhile, Szmrecsanyi defines a *linguistic corpus* as a methodology that refers to a more or less naturalistic collection of texts to perform linguistic analysis (Szmrecsanyi, 2017). McEnery and Hardie's corpus

linguistics is also defined as a set of studies in the form and function of language that combines computerized corpora in analysis. Corpus linguistics also focuses on extensive electronic data called corpora or data banks (Aarts et. ec, 2021). Regarding corpus linguistics, which is a language analysis method, this study uses collocation theory to analyze language data contained in the corpus. Therefore, this study aims to describe the collocation of netizen words in the 2022 Indonesian news contained in the Leipzig Corpora Collection.

LITERATURE REVIEW

Corpus linguistics

Corpus linguistics is a method of finding meaning from words and phrases to involve large amounts of text and the functioning of computer-assisted devices to analyze data (McEnery & Hardie, 2012). Corpus linguistics also focuses on procedures or methods of analyzing language-based software (McCarthy & A O'KeeFFE, 2010), which is the basic foundation of corpus linguistics focusesing on the study of language use. The observed data can be in the form of spoken and written data, which is processed, analyzed using corpus software and described using a linguistic approach, and described using the corpus software for specific points of view or goals (McCarthy & A O'KeeFFE, 2010).

Brazil (1995) defines a corpus as a collection of languages used. The meaning of 'languages spoken' is a language that has occurred in circumstances where the speaker is known to do something more than showing how the system works. This definition is helpful because it focuses on the fact that language in the corpus occurs naturally. However, we need to note that the corpus is not just a collection of languages that arise naturally in the form of isolated words or sentences that are collected randomly, but it consists of spoken and or written texts (the word text in corpus linguistics is used to refer to spoken and written language) (Timmis, 2015).

Collocation is a normal relationship between words in one language and certain other words (Umar, 2006). Meanwhile, Nattinger and DeCarrico (1992) define collocation as a series of words that seem to have the possibility that the sequence of words will appear together in any text.

On the other hand, Kridalaksana also stated that collocation is a continuous or permanent relationship and pairing of a word (in Hidayatullah, 2017). Meanwhile, according to Larson, collocation is a condition in which a word will appear in construction with other words (1984). Likewise Lea defines collocation as words that are combined in a language to produce pure writing (2002). In this case, Nida (in Arifin 2008) describes the types of collocation. According to Nida, there are two types of collocation, namely noun + adjective collocation (N + A) and adjective + noun collocation (A+N).

METHOD

This research is a corpus linguistics research because it uses the corpus as a tool of analysis. According to O'Keeffe and McCarthy Corpus Linguistics (Corpus Linguistics) is a method for analyzing language empirically by looking at language patterns from large data sets. The corpus used in this study is the Indonesian Corpus Leipzig (Yuliawati, 2014). Gries (2014) argues that corpus linguistics in the sense of words based on usage is the science of distribution. Quantitative linguistic corpus usually focuses on the frequency of occurrence through a list of frequencies, concordance, size of dispersion, and frequency of occurrence. All linguists aim at some form of generalization, but not all of them engage in some form of quantification to fulfill this goal. This means we can use the corpus both qualitatively and quantitatively. Qualitative corpus analysis may consist of formulating hypotheses and testing them by looking for potential rival samples in the corpus. It can also consist of using the corpus to refine the hypothesis at some stage (Desagulier, 2017)

The research method used in this study is a mix-method. According to Tashakori and Teddlie, mixed research is a general type of research in which the methods, techniques, or characteristics of the quantitative and qualitative paradigms are mixed in one overall study (in Yusuf, 2015). By using this, the researcher combines quantitative and qualitative analysis methods. Holmes & Meyerhoff argue that this combined method uses patterns identified through quantitative analysis as an essential basis for interpreting the text qualitatively in more detail (Holmes & Meyerhoff, 2010, p. 32). In relation to corpus linguistic research, Kwary and Arum stated that corpus linguistic research needs to use a mix-method because statistical data obtained from corpus analysis needs to be interpreted further using qualitative considerations (Kwary and Arum, 2011, p. 213).

The data for this research is the 2020 news contained in the Indonesian corpus Leipzig which contains 116,773,293 words, 959,268 types, and 7,335,309 sentences. The focus of the research is the collocation data of netizens which are coupled with adjectives. In this study, researchers used a web corpus called the Leipzig Corpora Collection. This website contains data from the 2020 Indonesian news corpus. This website can identify word patterns in the news published in the corpus by using the word search feature, then from the word, several sub-chapters will appear, namely 1) words with similar context, 2) examples, 3) words occurring in the same sentence, 4) neighbor co-occurrence left and right, and 5) word graph.

FINDINGS AND DISCUSSION

This study aimed to describe the collocation of netizen words in the 2020 Indonesian news corpus which the researchers took in the Leipzig Corpora Collection. Based on data from the Indonesian News Corpus in 2020, there were 116,773,293 tokens and 959,268 types. The frequency of occurrence of the word netizen in this corpus is as follows:

Table 1.1 the frequency of occurrence of the word netizen

Lexeme	Frequency of occurrence	Rank	Frequency class
Netizen	11,028	1,493	8

Based on the table above, the frequency of netizen in the Indonesian news corpus in 2020 amounted to 11,028. Then the netizen ranked 1,493, while the netizen occupied the class frequency at 8th place. In corpus analysis, if the class frequency is higher, the words have a lower frequency otherwise if a word has a lower class frequency, it means that the frequency of occurrence of the word has a higher frequency.

Referring to the focus of this research, the researcher found 2 collocation contexts in the Leipzig Corpora Collection which are side by side with the word netizen, namely the first Left Neighbour Cooccurrences (left context collocation) and Right Neighbour Cooccurrences (right context collocation). In the results of this discussion, the researcher divides the type of collocation of netizen into two sub-chapters which the researcher summarizes in the following table:

Table 1.2 the frequency of collocation of netizens' adjectives

Lexeme	Collocation context	Adjective	Frequency adjective
Netizen	Left	A lot	1,954
		A little	469
		Excited	309
		Angry	77
		Exclaim	66
		Wicked	52
		Negative	43
		Crowded	41
		Very funny	38
		Furious	28
		Suddenly	22
		Adorable	21
		Funny	11
		Shocked	6,4
		Curious	296
		Excited	229
		Adorable	182
		Envy	148
		Furious	127
	Right	Kagum	80
		Commotion	60
		Sarcastic	44

Surprised	30
Amazed	28
Suspect	16
Concerned	11
Worried	6,6

Based on the table, we can interpret that referring to the frequency of occurrence of left context adjective collocations in netizen word often collocates with many because the frequency of occurrence is 1,954, while the collocation of netizen is very little coupled with shocked because the frequency of its occurrence is only 6,4. Meanwhile, the collocation of right context adjectives in the word netizen often collocates with curious because the frequency of its occurrence is 296, while the collocation of netizen is very little coupled with worry because the frequency of its occurrence is only 6.6.

In connection with the findings of the data above, the researcher divides the adjective collocation of netizen into two sub-chapters, there are 1) adjective + noun collocation and 2) noun + adjective. Here is the presentation:

Collocation of noun + adjective

Nouns are a class of words that are addressed to objects or bodies or something animate that is special or general. While adjectives are a class of words that are usually added to nouns that indicate substance, to indicate quality or quantity (Kridalaksana, 2007). Noun + adjective collocation is a word pairing between a noun class that is side by side with an adjective. Collocation has two meanings, namely donotative meaning and metaphorical meaning (Arifin and Junaiyah, 2008). In this study, researchers found data relevant to collocation theory based on Nida's perspective. Below is the data on the collocation of netizen words in the 2020 Indonesian news contained in the leipzig corpus:

Table 1.3 Collocations of netizen words in the right context

Lexeme	Collocation type	Collocation context	Adjective	Frequency adjective
Netizen	<i>Noun+ adjective</i>	Right	Curious	296
			Up in arms	229
			Excited	182
			Envy	148
			Furious	127
			Impressed	80
			Commocotion	60
			Sarcastic	44
			Surprised	30
			Amazed	28

Suspect	16
Concerned	11
Worry	6,6

Referring to the table above, it is found that the collocation types of noun+adjectives in the Leipzig corpus are found along with their frequency of occurrence. Here is the presentation in detail:

The first type of researcher found the type of collocation of nouns (characterized by semantic + human, + concrete, + number) + adjective (characterized by semantic + character / action) contained in the Leipzig corpus. Here is the presentation:

- (1) Curious
- (2) Suspicious

The data above found that the word "curious and suspicious" collocated with netizens. Curious and suspicious are adjectives characterized by human semantics -concrete, spelled out, animate). The word "curious and suspicious" here is an adjective that shows a semantic character/action. On the other hand, the word "curious" has the meaning of wanting very much; or wanting to know, while the word "suspicious" has the meaning of being careful or cautious (because of worry, doubting, etc.), So from the meaning of the two it can be concluded that they are related to each other. In comparison, the netizen is a noun characterized by human semantics and concrete. When these two words are juxtaposed, they will form a new, specific meaning characteristic of netizens about the responses that netizens gave.

Second, the following data is also a collocation of the first type, namely collocation of noun + adjective characterized by character/action. The following is the explanation:

- (3) Envy
- (4) Sarcastic

The data above found that the words "envy and sarcastic" collocated with the word netizen. The words envious and julid are adjectives (adjectives) characterized by human semantics -, concrete). The words "envy and sarcastic" here are adjectives that show semantics characterized by character/action. On the other hand, the word "envy" the meaning of not being happy to see the advantages of others (lucky and so on, while the word "julid" has the meaning of envy and envy with the success of others, usually done by writing comments, statuses, or opinions on social media that corner certain people. Which of these meanings has something to do with each other By adjusting the existing circumstances, the nature of sarcastic itself often goes hand in hand with envy.

In comparison, the word netizen is a noun characterized by human semantics, concrete. If the two words are juxtaposed, it will form a new, specific meaning that is something that characterizes netizens.

Third, the following data is also the first type of collocation, namely noun + adjective collocation with character/action characteristics. The following is the explanation:

(5) Shocked

(6) Angry

The data above found that the words "shocked and angry" collocated with the word netizen. Words of surprise and anger are adjectives (adjectives) characterized by human semantics -, concrete). The word "surprised and angry" here is an adjective that shows a semantic character/action. On the other hand, the word "shocked" means a momentary emotional state caused by an unexpected event, while the word "angry" has an emotional meaning marked by conflict with someone or feelings after being treated wrongly. Which of each of these meanings has something to do with each other. By adjusting the existing circumstances, shock and anger are often associated with emotions or a person's reaction to a condition that is contrary to it. In comparison, the word netizen is a noun characterized by human semantics, concrete. When these two words are juxtaposed, they will form a new, specific meaning, which is something that characterizes netizens.

Fourth, the researcher found data on the collocation of nouns (characterized by -human semantics, -concrete, +countable, -group) + adjectives (characterized by +state) contained in the Leipzig corpus. Here is the presentation:

(7) Crowded

(8) Abuzz

The data above found that the word "crowded and abuzz" collocated with the word netizen. Crowding and busy are adjectives (adjectives) characterized by human semantics, concrete, and animate). The word "crowded and abuzz" here is an adjective that shows the semantic characteristics of the situation. On the other hand, the word "crowded" also has the meaning of being cheerful or festive. The word describes the state and gives rise to new activities. The word netizen itself is a noun characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Fifth, the researcher found data on the collocation of nouns (characterized by semantic -human, -concrete, +countable, -group) + adjective (characterized by +state) contained in the Leipzig corpus. Here is the presentation:

- (9) Suprised
- (10) Amazed
- (11) Impressed

The data above found that the words "Suprised, amazed and impressed" collocated with the word netizen. The words amazed, amazed and amazed are adjectives characterized by human semantics, concrete, and animate). The word "astonished and amazed" here is an adjective that shows the semantic characteristics of a character/deed. Which word describes the character of the response given by netizens to specific activities. On the other hand, the word "suprised" means feeling odd (when seeing or hearing something). While the word "amazed or impressed" means awe; wonder (of the greatness, beauty, beauty of someone or something). So when viewed, each meaning of the two words raises a relation to each other. In comparison, the word netizen itself is a noun that is characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Sixth, the researcher found data on the collocation of nouns (characterized by -human semantics, -concret, +uncountable, -group) + adjectives (characterized by +state) contained in the Leipzig corpus. Here is the presentation:

- (12) Excited
- (13) Commotion

The data above found that the word "excited and commotion" collocated with the word netizen. The words excited and uproar are adjectives (adjectives) characterized by human semantics, concrete, and animate). The word "excited and commotion" here is an adjective that shows semantics characterized by circumstances. On the other hand, the word "excited" means noisy, noisy, or uproar, while the word "commotion" means boisterous; uproar; up in arms; take it. Each of these meanings describes the similarity in terms of meaning because it provides a statement of certain feelings. In comparison, the word netizen itself is a noun that is characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Seventh, the researcher found data on the collocation of nouns (characterized by semantic -human, -concret, +countable, -group) + adjective (characterized by +self-state) contained in the Leipzig corpus. Here is the presentation:

- (14) Concerned
- (15) Worried

The data above found the word "concerned and worried" collocated with the word netizen. The words concerned and worried are adjectives characterized by human semantics, concrete, animate). The word "concerned and worried" here is an adjective that shows semantic characteristics of the situation. Which word describes the character of the response given by netizens to certain activities, this response is closely related to the word "concerned and worried." On the other hand, the word "concerned" has the meaning of being sad, worried, or indecisive (because of his failed efforts, getting into trouble, remembering his fate, and so on), while the word "worried" has the meaning of being afraid (anxious, anxious) about something that is not yet known, with certainty. Each of these meanings describes the similarity in terms of meaning because it provides a statement of certain feelings. In comparison, the word netizen itself is a noun that is characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Eighth, the researcher found data on the collocation of nouns (characterized by semantic -human, -concert, +countable, -group) + adjective (characterized by +self-state) contained in the Leipzig corpus. Here is the presentation:

- (16) Furious
- (17) Adorable

The data above found that the word "furious and adorable" collocated with the word netizen. Angry words are adjectives characterized by human semantics -, concrete, spelled out, animate). The word "furious and adorable" here is an adjective that shows a semantic character/action. On the other hand, the word "furious" also has the meaning of being very irritated or angry (inwardly), while the word "adorable" has the meaning of being very angry or exasperated, which are related to each other in terms of meaning and relation to their respective uses. From each of these meanings, it can be concluded that the word describes the situation that arises from activity and has a relationship with each other in terms of meaning. In comparison, the word netizen itself is a noun that is characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Adjective + noun collocation

Regarding collocations or word combinations, adjective + noun collocations are the opposite of noun + adjective collocations. This collocation is a word pairing between adjectives that are coupled with nouns. In this discussion, the researcher finds data relevant to the collocation theory based on Nida's perspective. The following is the data on the findings of collocations said by netizens in the 2020 Indonesian news contained in the Leipzig corpus.

Table 1.4 Collocation of left context netizen words

Lexeme	Collocation type	Collocation context	Adjective	Frequency adjective
Netizen	Adjective + Noun	Left	A lot	1,954
			A little	469
			Excited	309
			Angry	77
			Exclaim	66
			Wicked	52
			Negative	43
			Crowded	41
			Very funny	38
			Furious	28
			Suddenly	22
			Adorable	21
			Funny	11
			Shocked	6,4

Referring to the table above, it is found that the type of collocation of Adjective+ noun in the Leipzig corpus is found along with the frequency of their occurrence. Here is the presentation in detail:

Ninth, the researcher found collocation data with adjective (characterized + quantity) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is the presentation:

(18) A lot

(19) a little

The data above found that the words "a lot and a little" collocated with the word netizen. The words many and few are adjectives characterized by semantic count or quantity. On the other hand, the word "many" has a significant meaning in number, while the word "a little" has little meaning. When viewed from the meaning of each of these words, they are related to each other in terms of meaning and relation to their respective uses. In comparison, the word netizen is a noun characterized by human semantics, concrete. If the two words are juxtaposed, they will form a new, specific meaning, something that is a lot or a little from netizens.

Tenth, the researcher found collocation data with adjective (characterized + state) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is presentation:

(20) Excited

(21) Crowded

The data above found that the word "excited and crowded" is collocation with the word netizen. The word excited and crowded is an adjective (adjective) that is characterized by human semantics -, concrete, spelled out, animate). The word "excited and crowded" here is an adjective that shows the semantic characteristics of the situation. On the other hand, the word "heboh" means rowdy, scramble, or uproar. In comparison, the word "crowded" has the meaning of being cheerful or festive. When viewed from the meaning of each of these words, they are related to each other in terms of meaning and relation to their respective uses. In comparison, the word netizen is a noun characterized by human semantics and concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Eleventh, the researcher found collocation data with adjective (characterized + character/action + quality) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is the presentation:

(22)Wicked

(23)Negative

The data above found that the words "wicked and negative" collocated with the word netizen. Nasty and negative words are adjectives that are characterized by human semantics -, concrete, spelled out, animate). The word "evil and negative" here is an adjective that shows a semantic characteristic of character/action and quality. On the other hand, the word "evil" has a very ugly or bad meaning (about behavior, character, actions), while the word "negative" means not good or deviates from the general standard. When viewed from the meaning of each of these words, they are related to each other in terms of meaning and relation to their respective uses. In comparison, the word netizen is a noun characterized by human semantics and concrete. If the two words are juxtaposed, they will form a new, specific meaning, which is characteristic of netizens about the responses delivered on social media from netizens.

Twelfth, the researcher found collocation data with adjective (characterized + state) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is the presentation:

(24)Vey funny

(25)Funny

The data above found that the word "very funny and funny" collocated with the word netizen. Funny and funny words are adjectives (adjectives) characterized by human semantics -, concrete, spelled out, animate). The word "very funny and funny" here is an adjective that shows the semantic characteristics of the situation. On the other hand, the word "funny" means being ridiculous or witty, while the word "funny" means being funny or witty. When

viewed from the meaning of each of these words, they are related to each other in terms of meaning and relation to their respective uses. Meanwhile, the word netizen is a noun characterized by human semantics and concrete. If these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens about the responses delivered on social media from netizens.

Thirteenth, the researcher found collocation data with adjective (characterized + character/action) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is the presentation:

(26) Adorable

(27) furious

The data above found that the word "adorable and furious" collocated with the word netizen. Annoying and furious words are adjectives (adjectives) characterized by human semantics -, concrete, spelled out, animate). The word "annoyed and furious" here is an adjective that shows semantic character/action, where the word "adorable" has the meaning of being very irritated, or angry (in the heart), while the word "furious" has the meaning of being very angry or exasperated, which are related to each other in terms of meaning and relation to their respective uses. In comparison, the word netizen is a noun that is characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Fourteenth, the researcher found collocation data with adjective (semantic + character/action) + noun (semantic -human, -concert, +countable, -group) characteristics contained in the Leipzig corpus. Here is the presentation:

(28) Angry

(29) Exclaim

The data above found that the words "Exclaim and angry" collocated with the word netizen. Exclamations and anger are adjectives (adjectives) characterized by human semantics -, concrete, spelled out, animate). The word "exclaim and angry" here is an adjective that shows semantics characterized by character/action, where the word "exclaim" has the meaning of speech that is usually used for affirmation or high intonation, while the word "angry" has the meaning of anger. Because each of these meanings thus gives rise to a relationship with one another. In comparison, the word netizen is a noun characterized by human semantics, and concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

Fifteenth, the researcher found collocation data with adjective (characterized + state) + noun (characterized by semantic -human, -concert, +countable, -group) contained in the Leipzig corpus. Here is the presentation:

(30) Suddenly

(31) Shocked

The data above found that the word "suddenly and shocked" collocated with the word netizen. The words suddenly and shocked are adjectives (adjectives) characterized by human semantics -, concrete, spelled out, animate). The word "suddenly and shocked" here is an adjective that shows a semantic character/action. Which "suddenly" has an instantaneous meaning, which indicates a situation, while the word "surprised" means being surprised or surprised, which are related to each other in terms of meaning and relation to their respective uses. In comparison, the word netizen is a noun characterized by human semantics, concrete. When these two words are juxtaposed, they will form a specific new meaning, namely something characteristic of netizens.

CONCLUSION

Based on the explanation above, netizens said there were two collocation contexts: the first Left Neighbor Cooccurrence (collocation left context) and Right Neighbor Cooccurrence (proper collocation context). The types of collocations for netizens in the 2020 Indonesian news contained in the Leipzig corpus are 1) noun + adjective collocations consisting of the words curious, excited, excited, envious, furious, amazed, excited, julid, amazed, amazed, suspicious, concerned, and worried 2) Adjective + noun collocation consisting of many, few, excited, angry, exciting, evil, negative, crowded, funny, furious, suddenly, excited, funny, surprised.

Concerning collocation theory, collocation itself is still a hot topic because, as a result of collocation or word pairing, it can bring up two meanings, namely denotative meaning, and metaphorical meaning. Collocation in the corpus study is very closely related because of collocation. We can find out the word combinations that are often paired with a particular word or very rarely paired with that word. In the field of language translation, collocation is very important to pay attention to, to get the right and correct translation results according to the author's intent because collocation itself has many patterns, sometimes, there are common words and unusual words appear. In research, corpus-based analysis is used because the data contained in it is natural and empirical in that its form is actual linguistics.

There are still many shortcomings in this study due to the limitations of the researcher in classifying word classes and sorting out the data contained in the corpus due to a large number of words and variations. This is also because the corpus used is not self-made but uses an existing corpus, namely the Leipzig Corpora Collection. Seeing the number of corpus research, the researcher hopes

that further research can use a self-made corpus through the digitization process from a source to produce more valid, pure, and raw data from the language phenomena that occur.

REFERENCES

- Aarts, B., & McMahon, A. (2006). *The handbook of English linguistics*. Blackwell Publishing.
- Arifin, Z. J. (2008). *Sintaksis*. Jakarta. Grasindo
- Arlianti, D. N. (2020). Media Bagai Belati bermata Dua. In Adelia S. R. et al., (Eds) *Media Sosial, Identitas, Transformasi, dan Tantangannya*. Malang: Prodi Ilmu Komunikasi Universitas Muhammadiyah Malang & Inteligencia Media (Intrans Publishing Group).
- Cash, H., D Rae, C., H Steel, A., & Winkler, A. (2012). Internet addiction: A brief summary of research and practice. *Current psychiatry reviews*, 8(4), 292-298.
- Desagulier, G. (2017). *Corpus Linguistic and Statistic with R Introduction to Quantitative Methods In Linguistics*. Switzerland: Springer
- Goleman, D, (1999). *Working with Emotional Intelligence*. Gramedia Pustaka Utama,
- Hidayatullah, S. (2017). *Cakrawala Linguistik Arab*. Jakarta: Grasindo.
- Holmes, J. & Mayerhoff, M. (2003). *The Handbook of Language and Gender*. Blackwell Publishing.
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being?. *American psychologist*, 53(9), 1017.
- Kridalaksana, H. (2007). *Kelas Kata dalam Bahasa Indonesia*. Jakarta: Gramedia Pustaka Utama.
- Kwary, D.A & Arum, K.W.A. (2011). Lincoln's vs Obama's Presidencies: A diachronic corpus based analysis of the adjectival collocates of man and woman in the American English'. *Revel*, 9 (17), 211-225.

- Kwary D.A., Goldhahn, D., & Quasthoff, U. (2015). *Frequency dictionary: Indonesian*. Leipzig: Leipziger Universitätsverlag.
- Larson, M.L. (1984). *Meaning-based Translation: A Guide to Cross-Language Equivalence*. New York: University Press of America.
- Lea, D. (2002). *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- Leipzig Corpora Collection: Indonesian news corpus based on material from 2020. *Leipzig Corpora Collection*. Dataset. https://corpora.uni-leipzig.de?corpusId=ind_news_2020.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press,
- Nattinger, J.R. & DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. London: Longman
- O'Keeffe, A., & McCarthy, M. (2010) *The Handbook of Corpus Linguistics*. London: Routledge. London: Routledge, 2010.
- Rohmadi, A. (2016). *Tips Produktif Bersosial Media*. Jakarta: Elex Media Komputindo.
- Romli, A.S.M. (2018). *Jurnalistik Online*. Bandung: Nuasa Cendekia
- Sakti, M.N.S.F. (2020). *Moslem Social Media*. Jakarta: Elex Media Komputindo.
- Timmis, I. (2015). *Corpus Linguistics for ELT Research and Practice*. London: Routledge
- Umar, A.M. (2006). *Ilmu dalalah*. Kairo: Ilmu al-kutub.
- Yuliawati, S. (2014). Analisis berbasis Korpus: Kolokasi bermakna kata-kata "perempuan" dalam media Sunda (majalah Mangle, 2012-2013). *Jurnal Ranah*, 3(1), 107-123.
- Yusuf, M. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan Metode Gabungan*. Jakarta: Kencana