2019 Indonesia Presidential-Vice Presidential Debate in Corpus Linguistics Perspective

Susi Yuliawati[1]*, Eva Tuckyta Sari Sujatna[1], Dadang Suganda[1]
[1]Department of Linguistics, Faculty of Cultural Sciences, Universitas Padjadjaran, Indonesia
Corresponding Author: Susi Yuliawati, E-mail: susi.yuliawati@unpad.ac.id

ARTICLE INFO	ABSTRACT
  KEYWORDS *Corpus linguistics, keyword, lexical diversity, lexical word class, linguistic features, presidential debate, Indonesia*	The present paper carries out a corpus linguistic analysis of the first debate of 2019 Indonesia presidential election. The study compared the speech of presidential candidate pair number 1, JW and MA, and number 2, PS and SSU, in the debate in terms of lexical diversity and linguistic features. The research employs a mixed-method research design by using two corpus technical analyses, i.e. type/token ratio and keywords. Results show that PS spoke the most, while MA spoke with the most varied vocabulary. The study also found that foreign words from Arabic in the keywords of JW, PS, and SSU are generally used to show their belief as well as to embrace the Indonesian Muslim communities. However, the Arabic words used by MA tend to show his identity as a Muslim cleric, reflecting his in-depth understanding of Islam. Unlike the Arabic word usage, foreign words from English are used more for a practical reason and to emphasize arguments, particularly by JW and PS. Additionally, the study reveals that from the lexical word-class distribution, JW tends to give more focused information and entities, while PS tends to offer more explanations to present information. All things considered, the present writers argue that corpus linguistics is an essential method to investigate actual patterns of language used by politicians to interpret further. All in all, the present research is supposed to give a methodological contribution to the study discussing the relation between language and politics.

1. INTRODUCTION

The people of Indonesia enthusiastically witnessed one of the historic political events on 17 January 2019, i.e. the first debate of 2019 Indonesia presidential election, which was broadcast by a large number of media. In this political debate, we can perceive that language plays a significant role in politics. Aside from that, words in general also seem to be pivotal in Indonesia's political climate during the presidential election campaign. Such situation is not surprising given the notion that language shapes thought. The language that people speak even influences the most fundamental aspects of human experience and thus it is believed to be a determinant of reality (Borodistky, 2011). In today's digital era, information can be easily and rapidly accessed. As a result, not only valid information but also the misleading one can easily spread. Therefore, language is regarded as an instrument for power (Knappert, 2009) in political rivalries, particularly to persuade people to vote for them.

Hillary Clinton once reminded her opponent, Donald Trump, in the 2016 US presidential debate, saying

"*Words matter if you run for president*". In this case, Clinton assertively criticized Trump's off-the-cuff remarks and tweets, which had often been misleading, false, hateful, derogatory, inflammatory and juvenile. Trump, however, denied it by saying that he always delivered "*the best words*" (Gordon, 2017). However, to have confidence in their statements, it certainly needs to put them into an investigation, predominantly from the linguistic perspective. This speech event is evidence of how language has a substantial role in politics. By the language they use, politicians express their ideas and thought in knowledgeable and appropriate manners with the primary aim to induce people and get their vote. For this reason, we can find that some candidates of presidential debates or local leader election debates demonstrate oratory styles that may seem eloquent, elegant, calm or even sometimes forceful.

The rapid development of information technology has made language get more people's attention. In presidential debates, for instance, the competence and judiciousness of presidential candidates to project national interests are now possible to describe by the

language they speak. By using certain software to analyse their utterances, this can be expounded both qualitatively and quantitatively. From a transcript of a presidential debate, it can be identified, for instance, who has more turns, to what extent their responses are relevant and informative toward questions given by a moderator, and what words are frequently used and become the keywords of a candidate. A previous research on presidential debates is done by MS (2015) who studied some American presidential debates using corpus linguistics and functional grammar approach. He found that corpus linguistics provide crucial tools to identify the implications of selecting the lexico-grammatical tools, which are in turn crucial in enabling speakers to perform a number of functions such as constructing social relations, exercising power, and maintaining solidarity. The other research is from Chen, Yan, and Hu (2019) who investigated Clinton's and Trump's campaign speech during the general election by using corpus linguistics and discourse analysis approach. They found three major differences between Clinton's and Trump's linguistic styles as parts of their campaign strategy. Based on that, the present paper discusses the actual patterns of candidates' language use in the first debate of 2019 Indonesia presidential election by using a method of corpus linguistics. From the analysis of word frequency, in specific type/token ratio, and keywords, the study discusses the candidates' lexical diversity and linguistic features. This is expected to provide insight into how language is used to win the presidency-vice presidency from the opponents as well as to influence people to vote for them.

2. LITERATURE REVIEW

Corpus linguistics is considered rather different from the other branches of linguistics such as phonology, morphology, syntax, semantics, sociolinguistics, and pragmatics (McEnery & Hardie, 2012). In principle, research in language employing a corpus linguistics approach is associated with four major characteristics. First, the research is empirical with the aim to describe the actual patterns of language in use. Second, the research investigates a big and principled collection of natural text, known as corpus. The corpus designed and constructed represents a target domain of language use. Third, the research involves a far-reaching use of computer analysis employing either automatic or interactive techniques. Fourth, the research commonly integrates quantitative and qualitative analyses (Bibber & Reppen, 2015). In effect, corpus linguistics can be applied to study language from many different perspectives such as phonology, morphology syntax, semantics, pragmatics, and sociolinguistics.

Corpus linguistics has several distinctive analytical techniques, such as word frequency, keywords,

collocation, semantic preference, and semantic prosody. The approach regards that meaning of words are often created by the associations that the words participate in, alongside other words with which they frequently co-occurs, rather than by the words in isolation (Sinclair, 1991 & Stubbs, 2002). In this case, words tend to appear with certain words accompanying them in particular contexts, indicating the patterns of co-selected words that speakers and/or writers conform to (Sinclair, 1991). Thus, the approach considers meaning as a social construction (Yuliawati, 2018a). A corpus analysis to identify meaning based on this principle is known as the analysis of collocation. The term refers to a lexical relation between two or more words co-occurring within a few words of each other in running text. For example, the word PROVIDE frequently co-occurs with words referring to precious things that people need, such as *help* and *assistance, money, food* and *shelter,* and *information* (Cheng, 2012). In this case, the word PROVIDE is known as the node word, the word being investigated, while the words *help, assistance, money, food, shelter,* and *information* are called as the collocates, the co-occurring words in the corpus. Based on such collocational analysis, the word meaning can be examined.

Corpus linguistics is closely associated not only with qualitative analytical techniques but also quantitative analytical techniques to analyze real patterns of use in natural texts (Biber & Reppen, 2015). Therefore, corpus linguists often use statistical tests to obtain their quantitative findings. The most basic statistical test is word frequency analysis to identify words in a corpus, also known as tokens, from the most to the least frequent word. According to Cheng (2012), most corpus tools provide a program to generate word frequency lists easily and quickly and the lists are always interesting for further investigation. The lists include information about the total number of tokens in the corpus and also types, which are the number of unique or distinct words in a corpus. Besides, information about a type/token ratio (TTR), which can be used to measure lexical diversity, is often presented. According to Cheng (2012), a corpus with a high TTR indicates that it contains a diverse vocabulary or has few repetitions of words. In other words, the higher the TTR score, the greater the variety of types in the corpus.

The other typical corpus analysis is known as keyword analysis. In this case, keyword refers to lexical words that occur more frequently in a corpus under study (specialized corpus) than in a larger corpus (reference corpus) in which the difference in frequency is statistically significant (McEnery & Hardie, 2012). Therefore, keywords may signal the "aboutness" of

texts and are indicative not only of a discourse community but also of the writer/speaker's identity and position (Scott, 1997 and Bondi & Scott, 2010). Keywords also play an essential role in knowledge management, particularly to help researchers to pinpoint what items are worthy for further investigation in structured databases. With the help of corpus tools, keyword analysis is used to compare two lists of word frequency calculated using statistical metrics to accentuate interesting items which frequency differs significantly between one corpus that is being studied (specialized) and a much larger corpus (reference corpus). In general, the corpus tools generate keywords that are sorted by the keyness metric (usually using statistical significant test of chi-square or log-likelihood). This research used keywords analysis to identify some linguistic features that characterize the speech of the presidential and vice-presidential candidates in the debate. In this research, utterances of presidential candidates and their running mates in the first debate of 2019 Indonesia presidential election are examined by making use of the analysis of type/token ratio and keywords.

3. METHODOLOGY

The present study employs a mixed-method research design. By combining quantitative and qualitative analyses, the research is expected to gain a deeper understanding of 2019 Indonesia presidential-vice presidential debate. Mixed-method designs contribute to shed light on a better understanding of an object under investigation (Litosseliti, 2010). The quantitative approach is primarily used to collect comprehensive data and the qualitative approach is generally to interpret results of analysis (Yuliawati, 2018a).

For this research, a corpus of 2019 Indonesia presidential-vice presidential debate is constructed from the transcript of Indonesia presidential candidate debate of 2019 provided by an automatic transcriber machine, viz. NOTULA created by Bahasa Kita. The résumé of the transcript can be accessed from the website www.bahasakita.co.id. Bahasa Kita is an information technology company engaged in voice technologies with the specialization in the Indonesian language such as automatic speech recognition, voice biometrics, speech synthesis, speech identification, and natural language processing (NLP). The company has served various institutions including the Ministry of Defense of the Republic of Indonesia, the Corruption Eradication Commission, the Indonesian National Armed Forces, and Indonesia Deposit Insurance Corporation (IDIC).

Like the other corpus-based research, the present research utilizes a corpus tool for analysis, viz. Sketch Engine (Kilgarriff et al., 2014). The corpus tool provides several features to analyze language data such as wordlist, keyword, and collocation. This study makes use of these three features. Firstly, the feature of wordlist is operated to identify word frequency. Since the discussion includes the comparison of candidates' speech in terms of lexical diversity, the other corpus software, viz. WordSmith Tools (Scott, 2013), was used to generate type/token ratios. Secondly, the keyword feature is used to generate unique words, which are calculated based on keyness scores. The Sketch Engine provides the statistic formulation to generate the keyness score automatically, by comparing a specialized corpus, a corpus that is studied, with a reference corpus, a benchmark corpus which size is at least three times from the specialized corpus. For this research, the present writers built a corpus of Indonesia presidential-vice presidential debate consisting of 5,735 words. The corpus did not include moderators' utterances and thus the corpus was constructed from the utterances of four contestants in 2019 Indonesia presidential-vice presidential debate, i.e. Joko Widodo (JW), Ma'ruf Amin (MA), Prabowo Subianto (PS), and Sadiaga S. Uno (SSU). The corpus consists of 5,732 tokens. The reference corpus, which is used as the standard of comparison for the corpus of 2019 Indonesia presidential-vice presidential debate to generate keywords, is Indonesian web corpus (IndonesianWaC), a corpus of Indonesian language provided by the corpus tool *Sketch Engine* that consists of 90,120,046 words.

4. RESULTS AND DISCUSSION

The study that compares presidential and vice-presidential candidates' speech in the first debate of Indonesia presidential election 2019 using a corpus-based approach discusses two main topics. First, the analysis focuses on the lexical diversity of the presidential and vice-presidential candidates through the description of the frequency of token, type, and the ratio of type/token. Second, the study discusses the linguistic features in the corpus of 2019 Indonesia presidential-vice presidential debate by describing foreign word usages, lexical word-class distribution, and semantic categories of the candidates' keywords.

4.1 Lexical Diversity in 2019 Indonesia Presidential-Vice Presidential Candidates

As mentioned in the methodological section, the corpus of 2019 Indonesia presidential-vice presidential debate was built from the transcript provided by Bahasakita. Each of the candidates delivered utterances in the following numbers of tokens: 2,251 for JW corpus; 296 for MA corpus;

2,317 for PS corpus; and 871 for SSU corpus. Since the numbers of words are unequal, percentages and type/token ratios are needed to display a more accurate metric for the discussion on lexical profile comparison. The new percentages and type/token ratios, which were generated from the corpus software WordSmith Tools, are presented in the table below.

Table 1. Lexical profile of the candidates in 2019 Indonesia presidential-vice presidential debate

Corpus	Token	Type	Percentage of token	Type/token ratio
JW	2,251	723	39%	32.12
MA	296	177	5%	59.80
PS	2,317	770	41%	33.23
SSU	871	376	15%	43.17

As shown in Table 1, PS corpus has the highest percentage of token (41%). It suggests that among other candidates, PS is the one who delivered speech with the highest number of words in 2019 Indonesia presidential-vice presidential debate. Meanwhile, the type/token ratio is found to be the highest in MA corpus (59.80). According to Cheng (2012) the higher the TTR score, the greater the variety of types in the corpus. This indicates that MA's speech in the debate contains the most diverse vocabulary, in spite of the fact that the size of MA corpus is the smallest among the other candidates' corpus. In contrast, JW speech, which TTR score is 32.12, comprises the least diverse vocabulary, which means JW speech contains many repetitions of words although he spoke with the second-highest number of words in the debate.

According to Chen, Yu, & Han (2019, p. 19), it is generally believed that "politicians who speak in an accessible manner tend to be received by publics". JW corpus that is found to be the least in terms of lexical diversity apparently confirms this notion if it is related to the context of JW's victory in the April 17 presidential election, which was officially declared by Indonesia's General Elections Commission. In other words, JW spoke in a less complex language in terms of lexical diversity compared to PS' speech may have contributed to his victory. Meanwhile, corpus analysis showing that MA corpus size is the smallest is not surprising, since a lot of mass media and social media in Indonesia reported that MA spoke the least in the debate (Tehusijarana, 2019). However, as shown above, corpus linguistics approach can reveal not only the number of words spoken by the candidates in the debate but also their lexical diversity. Thus, we argue that the approach provides a methodological contribution to studies discussing the role of language in politics.

4.2 Linguistic Features in the corpus of 2019 Indonesia Presidential-Vice Presidential Debate

Linguistic features in this present article are identified through keyword analysis and interpreted in order to reveal linguistic styles that each of the presidential-vice presidential candidates presented in the debate, which may distinguish them. In principle, the analysis of keywords is used to determine which words characterize the text under investigation may be indicative of either what the text is about or what words are regarded essential (Yuliawati, 2018b). The extraction of keywords was processed utilizing the keyword module in the corpus software Sketch Engine. As stated in the methodological section, the procedure to generate the keywords is by comparing the word frequency list of the specialized corpus, which in this case is each of the presidential and vice-presidential candidates' corpora, and the word frequency list of the reference corpus, which is IndonesianWaC. The software generates the keywords ordered by the keyness scores. The following table presents the top 15 keywords in each of the candidates' corpora, ranked in descending order of keyness score.

Table 2. Top 15 keywords in the corpus of 2019 Indonesia presidential-vice presidential debate

No	JW	MA	PS	SSU
			Corpus	
1.	disabilitas	disabilitas	Jokowi	Sandy
2.	ASN	Jokowi	deradikalisasi	Prabowo
3.	calonkan	deradikalisasi	perkuat	partisipasi
4.	submission	ilal	permasalahkan	dipersekusi
5.	recruitment	intoleran	perbaiki	subhana
6.	Prabowo	radikalisasi	swatiastu	sinkronkan
7.	penindakan	mensinergikan	paslon	Cilamaya
8.	wathaniyah	perlakuannya	incorruptable	ASN
9.	paragame	terpapar	Buwas	Sandiaga
10.	jurkamnya	fasiq	tersakiti	difabel
11.	warohmatullah	keagamaannya	Bismillahirrah-maanirrahim	dikriminalisasi
12.	LHK	pendekatannya	masalahkan	ketidakber-hasilan
13.	melihat	khilaf	brightest	psikologinya

| 14. | *persilakan* | *diberantas* | *menatar* | *persekusi* |
| 15. | *pertentangkan* | *penindakan* | *ditangkep* | *radikalisasi* |

One of the noticeable linguistic features in the top 15 keywords of the corpus of 2019 Indonesia presidential-vice presidential debate is foreign word usage. Although all candidates' keywords contain foreign words, the composition is slightly different. As seen in Table 2, JW's and PS' unique words comprise foreign words from English and Arabic. JW mentioned three English words, i.e. *submission, recruitment,* and *paragame*, and two Arabic words, i.e. *wathaniyah*[1] and *warahmatullah*[2], while PS mentioned two English words, i.e. *incorruptible* and *brightest*, and one Arabic word, i.e. *Bismillahirrahmanirrahim* 'in the name of Allah, The Merciful, The Compassionate'. However, PS keywords includes not only foreign words from English and Arabic words but also Sanskrit, i.e. *swastiastu* 'customary Hindu greeting among Balinese people'. Unlike the presidential candidates, all vice-presidential candidates' unique words involve foreign words only from Arabic, i.e. MA used *ilal* (the word is from *ilal haq* meaning toward the truth) and *fasiq* (someone who violates Islamic law) and SSU mentioned *subhana* (the word is used for God in Islam from *subhanahu wataala*, meaning 'the most glorified, the most high').

According to Grosjean (in Kim, 2006), code switching, a language phenomenon when speakers switch or mix two languages, is often used as a communicative strategy to convey linguistic and social information. They use code switching, for instance, when they cannot find proper words or appropriate translation for the language being used. Additionally, Greene and Walker (2004) argue that code switching serves a function as a strategy at negotiation power to the speaker and it also reflects culture and identity, as well as promotes solidarity. It suggests that from the 15 top keywords, most of the candidates, particularly JW, PS, and SSU, used code mixing by inserting foreign words from Arabic, which are largely popular Islamic terms such as expression to greet, to begin something, and to say God, most likely not only to show their belief but also to embrace the people of Indonesia from Muslim community. In other words, code mixing is used here to maintain solidarity and also reflect identity. On the other hand, MA who used the Arabic word *ilal* and *fasiq* does not only expose his belief but also his identity as a Muslim cleric. The usage of Islamic term *ilal* and *fasiq* when

discussing programs to prevent Islamic radicalism indicates that he has an in-depth understanding of Islam. Unlike the Arabic word usage, Sanskrit word was only used by PS to maintain solidarity with the Hindu community of Indonesia because PS himself is a Muslim. In the meantime, the insertion of the English words in the candidates' speech is generally used when they could not find the proper words in Indonesian, or to emphasize their ideas.

The other interesting linguistic feature in the top 15 keywords of the corpus of 2019 Indonesia presidential-vice presidential debate to discuss is the lexical word-class distribution in each of the candidates' unique words. In the keywords of JW, MA, and SSU, the most dominant word-class is noun. According to Biber et al. (1999), nouns embody a high density of information. In line with that, Radford (2009) argued that nouns principally have semantic properties of denoting entities. The notions suggest that the high frequency of noun in the keywords of JW, MA, and SSU represents dense information and signifies a large number of entities, which were discussed in the debate. If we examine each of the candidates' keywords, it is also found that the highest percentage of nouns is in SSU's keywords (86.7%), followed by JW's keywords (80%) and MA's keywords (66.7%). In contrast to JW, MA, and SSU, PS used more verbs than nouns. The percentage of verbs in PS' keywords is 40% while the percentage of nouns is 33,3%. According to Radford (2009), verbs comprise the semantic property of signifying events and actions. As a result, the high percentage of verbs in PS's keyword may indicate that PS speech in the debate tends to contain more explanations about events and actions.

Furthermore, PS' keywords are also found to be the highest in the percentage of adjective, i.e. 13.3%. In the second position is found in MA's keywords, i.e. 6.6.7%, while JW and SSU do not have any adjective word-class in their top 15 keywords. However, the frequency of adjectives in the keyword is lower than nouns and verbs. Overall, the most dominant word-class in the 15 top keywords of the debate corpus is nouns and verbs are the second most dominant word-class in the Top 15 keywords in the corpus of the 2019 Indonesia presidential-vice presidential debate. If the two presidential candidates' keywords, JW and PS, are

---

[1] The word is from Islamic term *ukhuwah wathaniya* meaning maintaining mutual harmony among religious communities.

[2] The word is a part of the Islamic greeting *Assalamualaikum warahmatullahi wabarakatuh*, which means May the peace, mercy, and blessings of Allah be upon you.

examined further, it can be seen that their dominant word-lexical is different. JW's keywords contain more nouns than verbs. On the contrary, PS's keywords comprise more verbs than nouns. Based on the statement of Biber et al. (1999) as explain above, it may indicate that JW tends to present more focused information and entities, while PS tends to use more explanations to present information.

5. CONCLUSION

The present research has revealed that the candidates of 2019 Indonesia president-vice president presented similarities and differences in the way they delivered views and arguments in the first debate. All in all, PS corpus contains the highest number of tokens, indicating that he spoke the most in the debate. Then, based on the analysis of type/token ratio, MA has the highest score of TTR, while JW has the lowest score. It suggests that MA, the running mate of JW, used the most diverse vocabulary. On the contrary, JW spoke with the least varied vocabulary and this result apparently supports the general recognition that politicians who speak in accessible manner tend to be received by publics since JW has been declared officially to be the winner of the 2019 Indonesia presidential election.

Furthermore, the present study also reveals that foreign words are found in the top 15 keywords of all candidates who run for 2019 Indonesia presidential election. The foreign words used by them are from Arabic, English, and Sanskrit. However, each of the candidates used them for a slightly different purpose. Foreign words from Arabic found in the keywords of JW, PS, and SSU are apparently used to show their belief as well as to embrace the people of Indonesia from the Muslim community. Different from them, the Arabic words found in MA's keywords tend to reflect his identity as a Muslim cleric, showing his in-depth understanding of Islam. Unlike Arabic word usage, foreign words from English, found only in the top 15 keywords of JW and PS, are generally used to emphasize their ideas and for a practical reason.

From the lexical word-class distribution, it is also found that in the keywords of JW, MA, and SSU, the most dominant word-class is noun. Verb as the most frequent word is only discovered in PS' keywords. The result suggests that JW tends to give more focused information and entities, while PS tends to offer more explanations to present information.

5. ABOUT THE AUTHOR(S)

Susi Yuliawati is an assistant professor in Department of Linguistics Faculty of Cultural Sciences Universitas Padjadjaran, Indonesia. Her research interests are corpus linguistics, semantics, semiotics, language and gender.

Eva Tuckyta Sari Sujatna is an associate professor in Department of Linguistics Faculty of Cultural Sciences Universitas Padjadjaran, Indonesia. Her research interests are syntax, systemic functional linguistics, semantics, language and culture.

Dadang Suganda is a professor in Department of Linguistics Faculty of Cultural Sciences Universitas Padjadjaran, Indonesia. His research interests are pragmatics, language and culture.

REFERENCES

[1] Biber, D. & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics.* UK: Cambridge University Press.

[2] Bondi, M. & Scott, M. (2010). *Keyness in texts.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

[3] Borodistky, L. (2011). *How language shapes thought: The languages we speak affect our perceptions of the world.* [Online exclusive]. Scientific America. Retrieved from https://www.gwashingtonhs.org.

[4] Chen, X., Yan, Y., & Hu, J. (2019). A corpus-based study of Hillary Clinton's and Donald Trump's linguistic style. *International Journal of English Linguistics* 9(3), pp. 13–22.

[5] Cheng, W. (2012). *Exploring corpus linguistics: Language in action.* London & New York: Routledge.

[6] Gordon, J.K. (2017). *Words matter: Linguistic analysis of the presidential debate.* [Online exclusive]. Huffpost. Retrieved from https://www.huffingtonpost.com.

[7] Greene, M.D. & Walker, R.F. (2004). *Recommendations to public speaking instructors for the negotiation of code-switching practices among Black English-speaking African American students. The Journal of Negro Education* 73(4), pp. 435–442.

[8] Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *The Sketch Engine: Ten years on. Lexicography*, 1, p. 7–36.

[9] Kim, E. (2006). *Reasons and motivations for code mixing and code switching. Spring Issues in EFL* 4(1), pp. 43–61.

[10] Knappert, J. (2009). *The function of language in a political situation. Linguistics* 6(39), pp. 59–67. Retrieved from https://doi.org/10.1515/ling.1968.6.39.59

[11] Litosseliti, L. (2010). Research methods in linguistics. London & New York: Continuum.

[12] McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge, UK: Cambridge University Press

[13]   MS, El- Fakaly. (2015). *Vote for me! A corpus linguistic Analysis of American presidential debates using functional grammar. Arts and Social Sciences Journal* 6(4), pp. 1–13.

[14]   Radford, A. (2009). *An introduction to English sentence structure.* Cambridge: Cambridge University Press.

[15]   Scott, M. (2013). *WordSmith Tools manual: WS version 6.0*. Liverpool: Lexicaly Analysis Software Ltd.

[16]   Scott, M. (1997). *PC analysis of key words – and key key words. System,* 2 (2), pp. 233–245.

[17]   Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

[18]   Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishing.

[19]   Tehusijarana, K.M. (2019). *Ma'ruf Amin: Elderly cleric or canny political operator? We'll see Sunday.* [Online Exclusive]. The Jakarta Post. Retrieved from https://www.thejakartapost.com.

[20]   Yuliawati, S. (2018a). *Perempuan atau Wanita? Perbandingan berbasis korpus tentang leksikon berbias gender. Paradigma jurnal Kajian Budaya* 8(1), pp. 53–70.

[21]   Yuliawati, S. (2018b). *A corpus-based analysis of the terminology of the social sciences and humanities*. IOP Conference Series: Earth and Environmental Science, Vol. 175. http://doi.org/10.1088/1755-1315/175/1/012109