OPTIMIZING UAM CORPUS FOR TRANSITIVITY PROCESS REGARDING TO COVID-19 NEWS IN INDONESIA FROM FEBRUARY TO MARCH 2020

Heri Heryono
English Dept., Faculty of Languages, Widyatama University

Corresponding author: Heri Heryono, English Dept., Faculty of Languages, Widyatama University
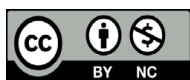E-mail: heri.heryono@widyatama.ac.id

Abstract:
In Systemic Functional Language (SFL), texts are analyzed by their transitivity system; and for the process are well-known as transitivity process. Practically, text parsing takes much time, since the part of text should be parsed manually and analyzed partially. By optimizing UAM corpus tool, the parsing process will get maximum results. UAM concerns on manual as well as semi-automatic annotation because of its lack of accuracy using automatic annotation. In this research, the texts are crawled from The Jakarta Post news article, especially in the form of online data. Those data (texts) are generated by UAM in order to get transitivity properties. The data are taken from February to March which contain Covid-19 issues in Indonesia. The result shows that between the time periods from February to March, the transitivity process regarding to the issue of Covid-19 in Indonesia has changed. It especially refers to the three process; material, mental and verbal process. During that period, material process slightly decreases from 58.72% to 57.59%, which emerges an assumption that people in Indonesia start to decrease their physical activity in March. While in mental process, it significantly elevates from 5.03% to 7.59% which could be assumed that people, including government, tend to think about the issue more seriously. While in the verbal process, it grows by about 2% from 16.11% in February to 18.62% in March which leads to an assumption that people have spread out the issue or the news to the other.

Keyword:
SFL, transitivity process, UAM corpus, Covid-19, news

INTRODUCTION

Corona virus case in Indonesia started out of control since it first came from several suspected patients in Jakarta last February. Spokesman handling of Covid-19 corona virus in Indonesia, Achmad Yurianto, said that the number of positive cases of the virus infection flourished by 21 cases to 117 people in recent month (March 2020). On February 14, 2020, patients infected with the corona virus danced with Japanese foreigners. This case has initially emerged as *the origin of Corona* in Indonesia which created a chaotic and panic situation from some people till the next month after (March). Based on the recent situation, there are three terminologies appeared in some social media between the people especially in Indonesia; *social distancing, panic buying* and *lockdown.* Those terminologies are found conveniently on *twitter, whatsapp* and another platform that share information regarding the Covid-19.

Drone Emprit has analyzed the complete data and released some insight and findings based on its analysis of conversation in social media and online media (Indonesian). Drone Emprit carried out the data collection on 25th of January 2020 when the issue of Covid-19 in Indonesia was considered as secondary proposition. By only one day, the news regarding Covid-19 slightly increased from one million mentions per-day to two millions mentions per-day. The data source is from Twitter account of @ismailfahmi.
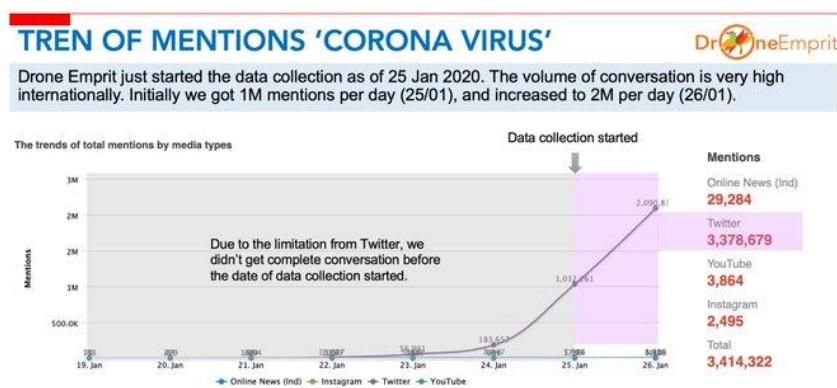
Image 1. Trend of mentions "Corona Virus"

Based on the report, Drone Emprit detected Indonesia becomes the country with the most conversations and followed by the other country, U.S and France by the hashtag #CoronavirusOutbreak.
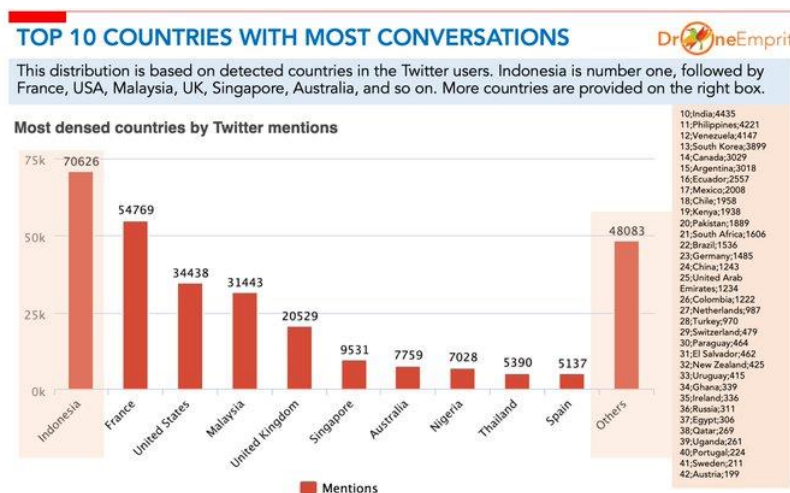

Image 2. Top 10 countries with most conversations

The data shows that the issue about Covid-19 increases daily, especially in Indonesia. It may be caused by the total numbers of active social users accessing from mobile device reach 130 million users (DataReportal, 2019). Government announcements regarding the positive case of the Corona virus against two Indonesian citizens have triggered panic among some people. The mass media was asked to preach rationally and not cause panic. An announcement from President Jokowi made people worried that they could contract the corona virus. In fact, some residents experience panic by shopping excessively or *panic buying*. The media must avoid news content that has the potential to trigger public panic. This noisy content is considered not to have a positive impact on the public or the government in handling Corona.

Information obtained from the news perched on human short-term brain memory (STM). However, news can also be a long-term memory if the news that is released is carried out with high frequency, then repeated with a high duration, so as the recent news related to Covid-19, especially in Indonesia. The context is different from what appears. The opponent's context is text (the text is not limited to writing but the meaning can be widened to whatever appears). Text provides denotative information, while context is an understanding of connotative text meaning. The connotative understanding of each person is different from the denotative meaning because it is influenced by cultural background, education, and any more.

Every media has an ideology in the news. The ideology of a media can be observed through language choices as well grammatical structure used in news writing. This ideology is usually adjusted to the target readers as news consumers. Ideological similarities between the media as well the reader is

considered important as part of the legitimacy of media domination towards the audience of readers (Eriyanto, 2000).

This research aims to identify online media effects by its relationship to SFL processes which are represented by *The Jakarta Post* on news coverage the disclosure of Covid-19 news in recent condition (February to March 2020 in Indonesia) against the after effect that influence people to think (mental process), to do something (material process) and to talk about that (verbal process). This research will be a significant endeavor to inform people about the alteration of paradigm regarding particular issue, in this term is Covid-19.

LITERATURE REVIEW
a. Transitivity Process

Systemic Functional Linguistics is the science of language (linguistics) which is also commonly called Functional Grammar. This kind of grammar is different from the previous (traditional) grammars which commonly regarded language as a set of rules of form. It looks at language with a broader perspective. Language is seen as a resource for making meaning (everything that allows us to be able to communicate). SFL is a system of meanings and other systems (systems of forms and expressions) to realize these meanings. This theory has two basic concepts, they are: a) language is a social phenomenon that is manifested as a social semiotic and b) language is a text that is construal (determines and refers to one another) with a social context. In relation to this study, LFS is more emphasized in terms of language functionalities according to the needs of the user (Halliday, 1985). In ideational meanings, language acts as the conveyor of ideas (that is, what is the essence of what someone is talking about, discussing, proposing, asking, etc.). So, in this first meta-function, the language we use every day certainly has an idea or subject. In it, more specifically we can review the choice of words used or other language features. The second meta-function is interpersonal meanings. From the term alone it may be guessed, interpersonal means the relationship between individuals, between people. In this meta-function, the main focus discussed is the function of language as a tool to bridge human relations. It can be analyzed of how close someone's relationship with the interlocutor just by analyzing the language used by using this 'tool'. The third meta-function is textual meanings, referring to how the message is conveyed. This can be seen from the organization or composition of information submitted, and whether the information is conveyed in oral or written medium (Thompson, 1991).

Transitivity is a system that describes experience as types of processes related to participants and circumferences. Transitivity is related to the choice of process type and the role of participants, which is realized in the reality of experience (Eggins, 2004). Transitivity can address how living things describe experiences based on the reality that is happening around both themselves and themselves. The aspects of experience are based on reality consists of: doing, happening, feeling, being. To illustrate the type of process selected in each clause, each related clause with different participant roles: actor, senser, behaver, sayer, extent, carrier. Whereas the types of processes involved in a clause consist of Material processes, Mental processes, Verbal processes, processes Behavioral, Relational processes, and Existential processes (Halliday, 1985)

First, the material process is related to treatment and shows what happened or event. The treatment is carried out by the body, physical, and action towards the material. A clause that uses a material process with action or treatment always with a real process or action. The main meaning of the material process shows that entities do people to something or what someone does to something. Syntactically this process can be stated in time (Thompson, 2014).

Second, mental processes are related to what is thought and felt. Halliday states the mental process as a process of feeling through 'feeling', 'perceiving', and 'thinking'. These three types are related to affection and reactive (feeling), cognition (thinking), senses (perceiving), or emotions or perception. Verbs are often used to say the mental processes of emotional or affection categories, for example, like and hate. Intellectual or cognitive categories, for example, believe, understand, and know. Taste categories through the five senses or perception such as looking at and hearing.

Third, the relational process is a process that shows the state (being). Every language in its grammar has a construction of a relational process. In English about three types of relational analysis. Fourth, behavioral processes which the process is characteristic between mental processes and

relational processes. Behavioral processes are physiological and psychological human behaviors such as *breathing, dreaming, smiling and staring.*

Fifth is the verbal process. Verbal process is often used in verbal language. This type of verbal process consists of three participants, namely the speaker (sayer), receiver, and words (verbiage). The speaker participant is responsible for the verbal process and participants.

Sixth is the existential process. This process uses there is / there was and there are / there were. This process shows the identification of the structure used by the word there. The use there has no particular meaning. That only shows the existential process. The important thing that distinguishes the use there as an existential subject and as a description of location / place. If the user there in the structure is usually not under pressure, while the use there in the information is usually under pressure and oriented.

Based on the discussion of the six processes above, transitivity is an important clause-forming element in the form of process. The six processes have their respective functions in the process of functional formation. The table below shows the summary of the meaning, participants and sample verbs in transitivity (Downing, 2006).

Table 1. Transitivity process

No	Processes	Category Meaning	Participants	Sample Verbs
1	Material			
		doing	Actor, Goal	*elect, give, choose*
		happening	Actor, Affected	*get, exposed*
2	Mental			
		perception	Senser-Phenonemon	*see, hear, notice, feel, taste, smell*
		affection	Senser-Phenonemon	*like, love, admire, miss, fear, hate*
		cognition	Senser-Phenonemon	*think, believe, know, doubt, remember, forget*
		volition	Senser-Phenonemon	*want, need, intend, desire, hope, wish*
3	Verbal	saying	Sayer, Receiver, Verbiage	*say, tell, pledge, express, address*
4	Relational			
		attributive	Carrier-Attribute	*to be, linking verbs: become, look, last*
		identifying	Token-Value	*to be, equal, signify, define*
		possessive	Possessor-Possessed	*have, has, possess, own*
5	Behavioral	behaving	Behaver	*smile, look, sniff*
6	Existential	existing	Existent	*to be*

b. UAM Corpus

This tool refers to a set of tools used in linguistic annotation, especially for texts which can be executed manually and semi-automatically. Later on, the application gives the user a media for searching texts for words or certain features, example: passive constructions and provides statistical analysis of the data. UAM tools basically focuses on manual and semi-automatic annotation, since its lack of accuracy using automatic annotation. Some linguistic pattern, as semantic or pragmatic features cannot be easily identified. In the other hand, appropriate language resources are required for well-analyzed automatic (O'Donnell, 2008).

To create a new project, the user has to provide a name at the preliminary process, for example, Text_Analysis_1 and it has to be specified the location of where the data will be saved. That project folder will contain various folders where the analysis, texts, schemes and results are arranged. The so-called project window (image 1) is the main window of the UAM Corpus Tool and helps to organize the data effectively.
CorpusTool stores its annotation data using XML. As with most modern annotation tools, it uses 'stand-off annotation' whereby the original text is left untouched, and annotation files refer to either character ranges in the original text, or ranges of tokens (Thompson and McKelvie 1997).
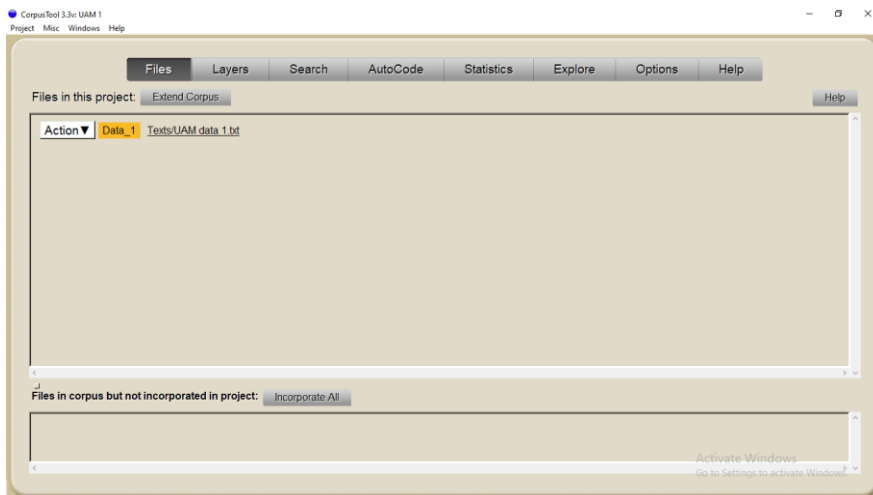
Image 3. Starting project window

After the window is opened by the user, the texts can be added by clicking the button *extend corpus*. The text should be in the plain format or *txt.* format. Currently, the only format that the users may apply is .txt only, but if the data should be in another format, it could be transferred or imported to be another format. As *word, excel* and *xml.*
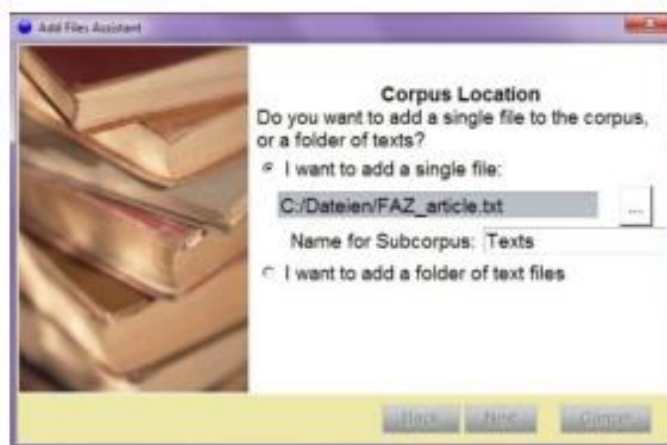

Image 4. Extend corpus

The next step is the metadata section where the user can edit described file. It is including language encoding, language information, and even the size of the font.

Table 2. Option

Button	Functionality
Change metadata	Edit language, encoding, font and font size.
View Basic Text Stats	Overview of general text statistics, e.g. the number of words and sentences, their average lengths, lexical density and the distribution of pronouns
Unincorporate file	Removes the file of your project, but it is still displayed as a file of your corpus, not completely deleted.
Clean MS chars	Removes special MS chars, e.g. unprintable chars.

The buttons on UAM Corpus have different functions as shown in the table above. When the researcher/user wants to edit language, encoding or change the fonts and sizes, user can use change metadata option. Moreover, the user can operate some particular buttons to get the corpus works properly, as text statistic function, lexical density and find out the word classes. This could

be the database for some research regarding to the words as the main data (as dataset), for example for sentiment analysis, SNA and for several linguistics research.
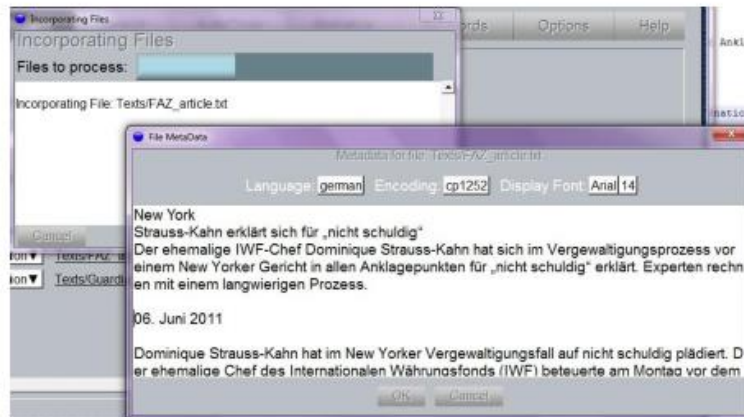

Image 5. File metadata

A graphical tools allow the users to generate a new scheme. Those are arranged hierarchically and the user may edit those by clicking on shown features (image 4).
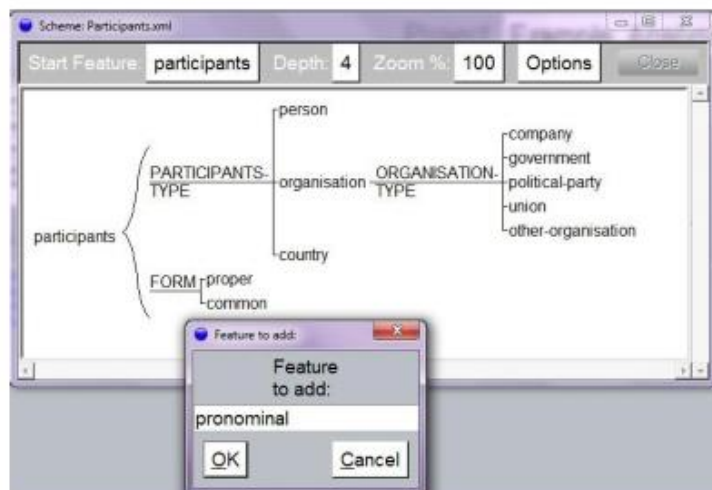

Image 6. Generating coding scheme

METHOD
The research simply used descriptive qualitative method in obtaining the result of the analysis. It included an intervention such as a training program (data training), some kind of social activity obtained from news articles and the introduction of a change in the person's living environment currently in pandemic situation. The research also applied a case study method, since the process of data collection and analysis were used. The data generated from The Jakarta Post articles which contained the issues related Covid-19 within first quarter of 2020. The researchers interested in a particular phenomenon of how people thought about the current situation regarding Covid-19. The method of this research was also categorized as non-participant observation research; and the researcher was not part of the group being studied (readers). The (raw) data in the form of texts accommodated sentence elements were processed as the dataset that subsequently became the main source to get the conclusion by generating UAM Corpus tool (Müller and Strube, 2006). In this case, corpora consisted of documents, abstracts, or sentences, and annotations that associated with the structured information (e.g. POS tags, named entities, shallow parses) with extents of the texts that could be analyzed (Orgen, 2006).
FINDINGS AND DISCUSSIONS

There are several steps before the researcher obtain the exact result by using UAM Corpus tool in analyzing texts from *The Jakarta Post* online news. The initial process is crawling the texts (news) from the (online) newspaper in the *.txt* format. Those articles were crawled by the period of February to March 2020 along with the Covid-19 massive reporting in Indonesia. The texts (news articles) taken from the newspaper were divided into two sections based on the month. By the result, the data set of texts consisted of *JakartaPostFeb.txt* and *JakartaPostMar.txt*. The file extension makes UAM corpus run effectively and fast, since the texts were clean. When the corpus were obtained from another media, by the example was twitter using API to crawl the dataset, the texts were full of trashes, as *hashtag, @,* RT, etc.
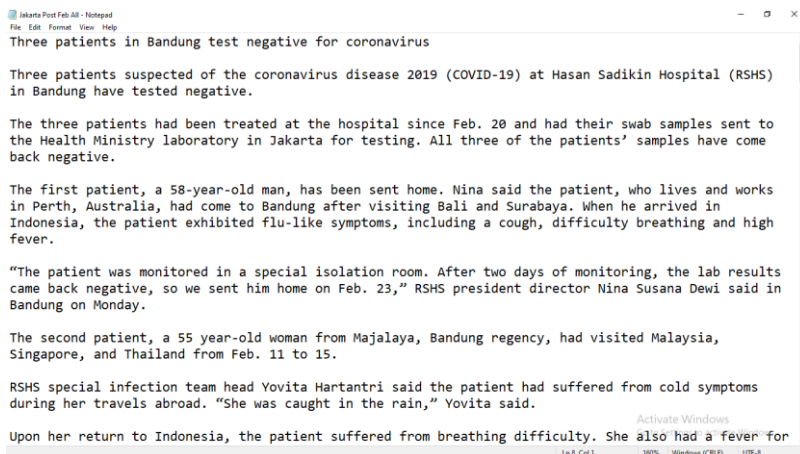

Image 7. Corpus in .txt format

After the datasets were crawled from *The Jakarta Post*, the next process was inputting the texts to UAM Corpus tools as previously been explained. The UAM tool would proceed the texts based on the command of the user. In this research, the texts were only focused on *transitivity* process which invloved the six processes.
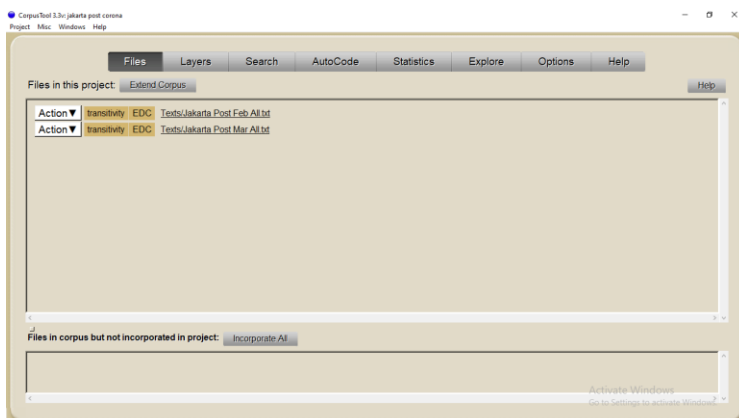

Image 8. Metadata process

Next step, after the texts were put in to UAM tool, the user might start to process or do the action in order to obtain the result. Click *transitivity* to complete the process, and then the UAM would self-processing until the results were shown in image 9.
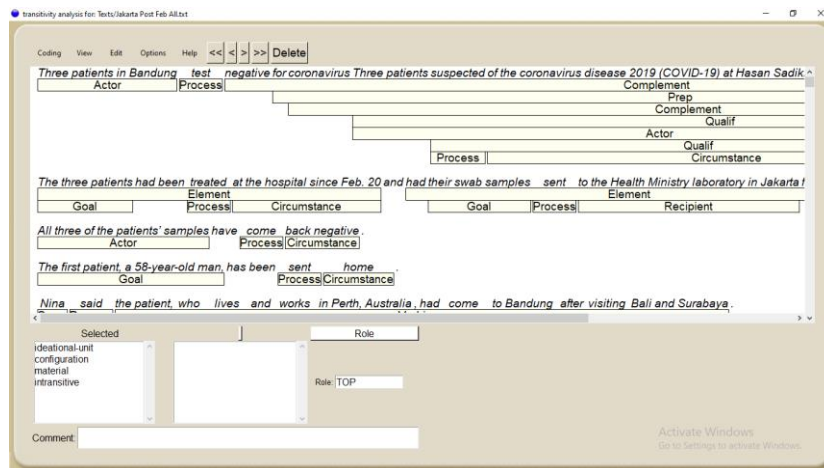
Image 9. Transitivity result

These elements of transitivity were clear and readable for the user. The hierarchical analysis could be seen in statistical section to get the result and to be imported to excel format.

Image 10. Transitivity process calculation in (%)

Based on the table above, it can be observed that the result of UAM corpus generating through the sentences resulting overview conclusions. The texts from source (The Jakarta Post) which contain specific issues of Covid-19 have different percentage (%) within selected months.

CONCLUSION
Types of transitivity processes in the news text analyzed consist of six types of processes, material processes, mental processes, relational processes, verbal processes, behavioral processes and existential process. In generating the transitivity process, the research uses UAM Corpus tools to obtain massive data set of clauses. By using the tool, each represented by a verb with each of its characteristics according to the type of transitivity process becomes more detail. The result of this research shows that between February to March, the transitivity process in the Covid-19 news in *The Jakarta Post* are slightly different. In this research, the focus is only to the three particular process; they are *material, mental* and *verbal process*. Based on the UAM Corpus analysis, in material process, the amount decreases from 58.72% to 57.59%, which leads to an assumption that people in Indonesia start to decrease their physical activity in March. This fact is supported by the condition of several institutions, government or private, to have *WFH (work from home)* system; even in some universities, they do the online learning to support the policy of Indonesian government. While in mental process, it significantly increases from 5.03% to 7.59% which could be assumed that people, including government, tend to think about the issue more seriously. It is marked by starting up some (mobile) applications, building up new health-care systems, creating new regulation (by applying standard procedure for sanitizing in several public facilities) and even

arranging *lockdown* regulation. While in the verbal process, it flourishes by about 2% from 16.11% in February to 18.62% in March which leads to an assumption that people have spread out the issue or the news to the other. Even hoax emerged in every social media platform, especially about the *lockdown policy* in March 2020 regarding the Covid-19 outbreak. At this point, using UAM corpus tool to identify and analyze the news is effective especially in very large amount of data set (news article). It leads to the bigger output that may be obtained by researcher to get conclusion for the researcher.

REFERENCES

Downing, Angela and Locke, Philip. (2006). *English Grammar: A University Course: 2nd Edition.* New York: Routledge, 2006. Print.

Eggins, Suzanne. (2004). *An Introduction to Systemic Functional Linguistics: 2nd Edition.* London: MPG Books Ltd, Bodmin, Cornwall, 2004. Print.

Eriyanto. (2000). *Analisis Wacana.* Malang: LkiS.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar.* London: Edward Arnold.

https://twitter.com/ismailfahmi/status/1221505763450875904

https://www.slideshare.net/DataReportal/digital-2019-indonesia-january-2019-v01

Müller, C. and M. Strube 2006. 'Multi-Level Annotation of Linguistic Data with MMAX2', in S. Braun, K. Kohn, J. Mukherjee (eds.) Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods (English Corpus Linguistics, Vol.3). Frankfurt: Peter Lang. 197-214.

Ogren, P.V. 2006. 'Knowtator: a plug-in for creating training and evaluation data sets for biomedical natural language systems', Proceedings of the 9th International Protégé Conference. 73–76.

O'Donnell, M. (2008). *Demonstration of the UAM CorpusTool for text and image annotation.* In Proceedings of the ACL-08: HLT Demo Session (Companion Volume), pp. 13-16. Association for Computational Linguistics. (last access 16.10.2011)

Thompson, Geoff. (1991). *Introducing Functional Grammar.* China: Edward Arnold (Publishers) Ltd.

Thompson, Geoff. (2014). *Introducing to Functional Grammar Third Edition.* London: Routledge.

Thompson, H. and D. McKelvie. (1997). *Hyperlink semantics for standoff markup of read-only documents.* Proceedings of SGML Europe'97.