Lexical Richness of *the Jakarta Post* Opinion Articles: Comparison between Native and Non-Native Writers

Dwi Indarti
Akademi Bahasa Asing BSI Jakarta
Jl. Salemba Tengah No 45 Jakarta Pusat
Email: dwi.diw@bsi.ac.id

Abstract – *Tujuan dari penelitian ini adalah untuk mencari kompleksitas perbendaharaan kata-kata yang ada di dalam artikel opini surat kabar. Dengan menggunakan dua buah artikel opini yang mengangkat topik yang sama mengenai President Donald Trump yang terbit di koran* The Jakarta Post *dan ditulis oleh seorang penutur asli bahasa Inggris dan seorang yang bukan penutur asli bahasa Inggris, penelitian ini membandingkan tingkat kompleksitas perbendaharaan kata antara kedua penulis. Kedua buah artikel tersebut adalah* **Trump's muslim travel ban will spark radicalization** *yang ditulis oleh Moh. Zaki Arrobi dari Indonesia dan* **Trump is attacking religious liberty** *yang ditulis oleh Noah Feldman dari United Kingdom. Penelitian ini menggunakan sebuah program bahasa berbasis web melalui situs* www.lextutor.ca *yang bisa diakses secara bebas, bernama* Lexical Frequency Profile *yang digagas oleh Laufer and Nation (1995). Hasil penelitian ini membuktikan pendapat Laufer and Nation (1995) yang menyatakan bahwa penulis yang bukan penutur asli bahasa Inggris mempunyai perbendaharaan kata yang lebih sedikit dibanding dengan penulis yang merupakan penutur asli bahasa Inggris. Penelitian ini menunjukkan adanya perbedaan yang cukup penting.*

Kata kunci: *Kompleksitas perbendaharaan kata, Artikel opini surat kabar, Penutur asli Bahasa Inggris, Penutur Tidak Asli Bahasa Inggris*

I. INTRODUCTION

Newspapers are daily paper publications contain information about the current news and related articles. They are divided into some sections such as news articles, editorials articles, features articles and opinion articles. Masroor (2013, p. 35) stated that, "An important movement of newspaper industry is the separation of news pages from public opinion. The movement not only gave birth to the notion of impartial journalism, but also authorized newspapers in the explicit articulation of opinions on issues of contention through opinion discourse." Newspapers give the opportunity to the readers to convey their thoughts and ideas on some particular topics based on their point of views.

Opinion articles are short argumentative essays, usually around 2000-2500 words, written by the newspaper's readers. They are different from editorial articles as Saleh (2013, p. 211-212) explained that, "The difference between editorial and opinion article is editorial articles represent the views of publication and their content is dictated by the publisher or an editorial board while opinion articles represent opinions of authors neither associated with the editorial board nor the publication itself" Opinion articles are opened for society to express their ideas, thought and opinion on current and popular issues such as politics, economy, finance, human rights, technology, culture, education and other interesting topics.

As an essay, an opinion article consists of four main structures; the title, the opening, the body and the conclusion paragraphs. The title is to attract the readers to read the whole essay, while the introduction paragraph presents a clear thesis statement of the whole article. Meanwhile, the body paragraphs provides supporting details, specific evidences, data statistics and examples and conclusion paragraph includes the brief summary of the whole idea.

The Jakarta Post is a daily English newspaper published in Indonesia and provides specific pages for opinion articles written by both native and non-native writers. The writers of opinion articles in *The Jakarta Post* are from native countries (USA, UK, Australia), English as a Second Language (ESL) countries such as Philippine, India, French, and so on), and English as a Foreign countries (Indonesia, Japan, China, Korea, UEA, and many more). Each edition contains 5 to 6 opinion articles with various topics related to the current issues.

Studies of opinion articles have been flourishing in linguistic area. Some studies related to opinion articles have been conducted by scholars. Ansary and Babaii (2009) used the opinion articles published in *The Washington Time, The Iran News*, and *The Pakistan Today* to characterized the global and/or macro-rhetorical structure of English newspaper opinion articles with Generic Structure Potential (GSP). Hoffman and Slater (2007) explored Schwartz's value framework to characterized lay and journalist frames in a national sample of local newspaper. The study applied a content analysis on newspaper opinion articles on health policy issues and revealed that there were

significant differences in values were found by the article type and topic.

Another remarkable paper uses newspaper discourse is Milne (2008) who conducted a cross-linguistics study about the pragmatic role of textual and interpersonal meta-discourse markers in the construction and attainment of persuasion. Using a corpus of 40 opinion articles, 20 written in English are taken from the British *The Times* and 20 written in Spannish are taken from *El Pai*, the study purposed to identify the categories of meta-discourse that predominate the newspaper opinion articles.

Meanwhile, in Indonesian context, Pulungan, Subroto, Tarjana and Sumarlam (2010) investigated the inter-textuality of Indonesian newspaper opinion articles on education. Using six newspapers published in some areas of Indonesia (*Kompas, Waspada, Pikiran Rakyat, Kedulatan Rakyat, Suara Merdeka, Fajar*), their study revealed that there are three major types and two major functions of inter-textuality in the opinion articles.

The quality of an essay can be seen from its lexical richness (Laufer and Nation : 1995). Lexical richness is the variety of vocabulary size used by an author to write an essay. "Lexical richness is the result of learning new vocabulary, activating previously known vocabulary or proficient in the writing skill" (Laufer and Nation, 1995, p. 308).

Furthermore, Laufer and Nation (1995, p. 307) stated that," a well-written composition, among other things, makes effective use of vocabulary. This need not be reflected in a rich vocabulary, but a well-used rich vocabulary is likely to have a positive effect on the reader.". Measures of lexical richness attempt to quantify the degree to which a writer is using a varied and large vocabulary. (Laufer and Nation, 1995, p. 307).

There are several measures of lexical richness proposed by Laufer and Nation (1995); Lexical Originality (LO), Lexical Density (LD), Lexical Sophistication (LS), Lexical Variation (LV) and Lexical Frequency Profile (LFP).

1. Lexical Originality

It is the percentage of words in a given piece of writing that are used by one particular writer and no one else in the group.

$$LO = \frac{Number\ of\ tokens\ unique\ to\ one\ writer\ x\ 100}{Total\ number\ of\ tokens}$$

The Lexical Originality index measures the learners' performance relative to the group in which the composition was written. (Laufer and Nation, 1995, p. 309).

2. Lexical Density

It is defined as the percentage of lexical words in the text, such as nouns, verbs, adjectives, and adverbs.

$$LD = \frac{Number\ of\ lexical\ tokens\ x\ 100}{Total\ number\ of\ tokens}$$

Lexical Density does not necessarily measure lexis, since it depends on the syntactic and cohesive properties of the composition. (Laufer and Nation, 1995, p. 309).

3. Lexical Sophistication

It is the percentage of 'advanced' words in the text.

$$LS = \frac{Number\ of\ advanced\ tokens\ x\ 100}{Total\ number\ of\ lexical\ tokens}$$

What is labeled as 'advanced' would depend on the researchers' definition. To decide what vocabulary is advanced, it is necessary to take the learner's level into consideration. Lexical Sophistication is determined by the researcher's definition of advanced or sophisticated words, its uses are limited. (Laufer and Nation, 1995, p. 309-310).

4. Lexical Variation

It is the type/token ratio in per cent between the different words in the text and the total number of running words.

$$LV = \frac{Number\ of\ types\ x\ 100}{Number\ of\ tokens}$$

The type/token ratio has been shown to be unstable for short texts and can be affected by differences in text length; even more sophisticated formulas have been shown to be unsuitable for short texts like learners' essays. LV distinguishes only between the different words used in a composition. (Laufer and Nation, 1995, p. 310).

5. Lexical Frequency Profile

The LFP (Lexical Frequency Profile) shows the percentage of words at different vocabulary frequency levels. The calculation is done by a computer program which compares vocabulary lists against a text that has been typed to see what words in the text are and are not in the list and to see what percentage of the items in the text are covered by the lists. The program can calculate the LFP on the basis of word tokens, word types or word families. (Laufer and Nation, 1995, p. 312). The classification of words is based on the General Service List (GSL) Published by Michael West (1953) which represents the most frequent words of English. It consists of K1 words or one thousand most frequent words of English (1-1000) and K2

words or two thousand most frequent words of English (1001-2000). The program also shows the AWL (Academic Word List) that is words with high frequency appearance in English academic texts. The AWL was compiled by Averil Coxhead from the Victoria University of Wellington and contains of 570 word families.

Below is the terminology of lexical richness

Table 1. Terminology of lexical richness

The First 1000 words (1-1000)	K1 words	
The Second 1000 words (1001-2000)	K2 words	
Academic Word Lists	AWL	
Off-List words	-	
Number of words	Tokens	N
Number of different words	Types	V
Number of words occurring once	V_1	V_1
Type/token ratio	TTR	V/N

"Token" is the total number of words occurring in a text, while "Type" refers to the number of distinct words in a text. For example, the sentence *A day without a book is like a house without a window* has 12 tokens and 8 types (a and without are repeated).

Type/token ratio (TTR) is the number of types divided by the number of token. The TTR of the sentence *A day without a book is like a house without a window* is $\frac{8\ (types)}{12\ (tokens)} = 0,6\ or\ 67\%$

If the writer uses many the same words (same types), then the TTR score will be low. It means that the text does not have very lexical richness. A big TTR score indicates a high lexical richness, while a small TTR shows a low lexical richness.

Therefore, this study aims to measure the quality of *The Jakarta Post* opinion articles written by native and non-native authors by investigating its lexical richness. It is also of interest to reveal whether there is a difference between native and non-native writers in term of lexical richness. This study uses the *Lexical Frequency Profile* proposed by Laufer and Nation (1995) to reveal the data of lexical richness of the opinion articles from *The Jakarta Post*.

II. RESEARCH METHODOLOGY

Research Paradigm

T This study used quantitative method since the data is numerical. Woodwars (1930) as cited in Hamilton and Lawrence (2012, p.6) stated that,"Quantitave analysis is one of the possible ways to conduct research in news, especially in public opinion articles. Where some statistical and semi-behaviorist of data can be used to the study of social phenomenon called public opinion." It means that since the result of this study shows a list of numbers, then the appropriate method to analyze it is to use the quantitative method.

The Source of Data

This study took two opinion articles from *The Jakarta Post*. Those two opinion articles were published in April 2017 edition and discussed the same topic about Donald Trump's policy. The details of data used in this study are described as follow:

Native Text : *Trump is attacking religious liberty*

Author : Noah Feldman

About the author : The writer is a Bloomberg view columnist. He is a professor of constitutional and international law at Harvard University and was a clerk to US Supreme Court Justice David Souter. His books include Cool War: The Future Global Competition and Divided by God: America's Chruch-State Problem and What We Should Do About It.

Non-native text : *Trump's Muslim travel ban will spark radicalization*

Author : Moh Zaki Arrobi

About the author: The writer is an Indonesian postgraduate student at the University of Essex, England.

Technique of Data Collection and Analysis

There were several steps taken in collecting and analyzing the data. First, the write searched the opinion articles with the same topic that published in the same edition. The next step was to re-type the printed opinion articles in Microsoft Words program before they were processed using the free-web computerized online digital tools via www.lextutor.ca. This free-web computerized program provides the Lexical Frequency Profile. The user choose the 'vocabulary Profile' menu then 'VP-Classic' menu and the next is just have to input all the text into the program, finished by press the 'SUBMIT_Window' button.

III. RESULT AND DISCUSSION

After passed through all the process, the writer got the result as showed by the tables below:

NON-NATIVE TEXT

Table 2. Non-native text

	Families	Types	Tokens	%
K1 Words	179	204	580	68.32%
K2 Words	22	26	35	4.12%
AWL Words	42	47	66	7.77%
Off-List Words		103	168	19.79%

Table 2 above shows that in non-native text, the writer found the families of K1 words are 179, K2 words are 22, AWL words are 42. Non-native text contains types of K1 Words amount 204, K2 Words amount 26, AWL Words amount 47 and Off-list Words amount 103. Meanwhile, tokens of K1 words are 580, tokens of K2 words are 35, tokens of AWL are 66. Off-list words are words that do not belong to K1, K2 and AWL. Those words include the name of persons, places, local language words, such as Mohammad, Afganistan, Irak, Indonesia, and so on. There are 42 words categorized as Academic Word Lists (AWL). Those words are *academy, acknowledge, administer, benefit, classic, community, consequent, contrast, controversy, error, exploit, file, furthermore, generate, globe, individual, input, instruct, liberal, major, maximize, minor, obvious, option, paradigm, perceive, policy, pose, potential, previous, prohibit, publication, radical, regime, research, secure, submit, sum, target, text, trigger, widespread.* The whole words in text (tokens) found in non-native texts are 849 with different words (types) amount 380. The type token ratio (TTR) is 0.45.

NATIVE TEXT

Table 3 Native text

	Families	Types	Tokens	%
K1 Words	218	266	882	75.97%
K2 Words	29	42	54	4.65%
AWL Words	51	61	96	8.27%
Off-List Words		83	129	11.11%

Table 3 shows that in native text, the writer found the families of K1 words are 218, the families of K2 words are 29, the families of AWL words are 51. Meanwhile, the types of K1 words are 266, the type of K2 words are 42, the type of AWL words are 61, and the type of off-list words are 83. Tokens of K1 words are 882, tokens of K2 words are 54, tokens of AWL words are 96 and tokens of off-list words are 129.

There are 51 words categorized as Academic Word Lists (AWL). Those words are *academic, access, affects, amendment, analogous, authority, behalf,*

challenge, clarified, classic, clause, constitution, constitutional, definition, denied, designed, dimensions, discrimination, display, errors, establishment, files, global, goal, guarantees, immigrants, immigration, injured, input, instance, instructions, interpreted, legal, majority, maximum, minorities, network, neutral, option, physically, prioritization, prohibits, research, security, status, submit, symbolic, targeting, text, theory, tradition, valid, violates.

Table 4. Comparison Between Native And Non-Native Text

NATIVE	NATIVE	NON-NATIVE
Words in text (tokens)	1.032	681
Different words (types)	369	277
Families	308	243
Type-token ratio (TTR)	3.35	2.80
Academic Word Lists (AWL)	51	42

From Table 4, we can see that there is a quite significant different between native and non-native text. Native text has 1.032 tokens, while non-native text has 681 tokens. It means that native writer uses more vocabularies in expressing his idea about the topic. On the other hand, non-native writer uses less vocabulary in sharing his opinion about the issue. Native writer has more different words (types) and families (369) and (308) compare to non-native writer's (277) and (243). It shows that native writer explores more vocabularies than non-native writer in writing the opinion articles in newspaper. Type-token ratio (TTR) shows that native writer has higher score (3.35) than non-native writer (2.80). In term on Academic Word Lists (AWL), native writer uses more academic words (51) than non-native writer (42). The result indicates that native writer has more rich vocabularies than non-native writer. This result supports the argumentation proposed by Laufer and Nation (1995) who stated that non-native writers have a relatively small vocabulary size compared with native writers. It could be understable since native writer uses English as the mother tongue, while non-native writer considers English as a second language or a foreign language.

IV. CONCLUSION

This study aimed to seek the variety of the vocabulary size used by authors of *The Jakarta Post*'s opinion articles. Using two opinion articles with the same topic which published in *The Jakarta Post* in April 2017 edition, this study compared the

lexical richness of native and non-native writers. This study followed the formula to measure the lexical richness called Lexical Frequency Profile, proposed by Laufer and Nation (1995) that can be accessed freely through www.lextutor.ca.

The result of this study showed that Laufer and Nation's (1995) argumentation is proofed. Native writer had higher Type-Token Ratio (TTR)'s score than non-native writer. The high TTR score indicated that the writer explored more variety of vocabularies in expressing his ideas of specific topic of writing.

This study is expected to give contribution to the linguistic field, specifically in professional discourse genre such as newspaper's opinion articles. For the future research, it is of interest to reveal other linguistic aspects, such as syntactical complexity and lexical density of a text, such as editorials or news pages.

REFERENCE

Ansary, H., & Babaii, E. (2009). A cross-cultural analysis of English newspaper editorials: A systemic-function view of text for contrastive rhetoric research. *Regional Language Center Journal, 40*(2), 211-249.

Hamilton, J., & Lawrence, R. (2012). *Journalism studies: Theory and practice*. Taylor & Francis: New York.

Hoffman, L.H., & Slater, M.D. (2007). Evaluating public discourse in Newspaper opinion articles: Values-framing and integrative complexity in substance and health policy issues. *Journalism & Mass Communication Quaterly, 84*(1).

Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistic, 16*(3), 307-322.

Masroor, F. (2003). Argumentative strategies on newspaper editorials in English across cultures. *Asian ESP Journal, 9*(3), 35-72

Milne, E. D. (2008). The pragmatic role of textual and interpersonal meta-discourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics, 40*, 95-113.

Saleh, B. (2013). The complete guide to article writing: How to write successful articles for online and print markets. *Writer's Diggest Books*: Ohio.