

Lexical Richness of Short Stories Written by EFL Students

Nata Permata¹, Rosyida Ekawati²

^{1,2}Universitas Trunojoyo Madura

Email: permataagdiar@gmail.com

Abstract

The study concerns with lexical richness of short stories. The data consist of 48 short stories written by English Study Program of University of Trunojoyo Madura. This study aims at measuring the lexical richness of short stories by different gender based on three measurements, i.e. lexical density, lexical sophistication, and lexical variation. This study employs the theory and the lexical richness formula proposed by Laufer and Nation (1995). It applies descriptive quantitative design. The results showed that the results of each measurements are 11 short stories in high lexical density (LD), 37 short stories in quite high LD, 27 short stories in high Lexical Sophistication (LS), 21 short stories in low LS, 25 short stories in high Lexical Variation (LV), and 23 short stories in low LV, the results of lexical richness that obtained from the combination of LD, LS, and LV are high lexical richness category (5 stories), quite high category (14 stories), quite low category (20 stories), and low category (9 stories), and the results of lexical richness based gender differences show that the short stories of male authors have higher lexical richness than female authors. The overall results show that short stories written by English Study Program students of University of Trunojoyo Madura have quite low lexical richness.

Keywords: Lexical richness, lexical sophistication, lexical variation,

Introduction

English can be said as a global language because the use of English is recognized in every country and it plays the special role. According to Crystal (2003), more than 100 countries use English as the main foreign language that should be encountered in schools, replacing other foreign languages. This role is seen obviously in some countries which do not use English as a mother tongue. In Indonesia, English is considered to be foreign language. Foreign language refers to the language that is not spoken in the society but it is learned largely in the classroom (Moeller & Catalano, 2015). Since English is a foreign language, it is taught from elementary school student until adult. However, it is not required for the society in Indonesia to be able to master. Even so, English is important to be learned for supporting someone's knowledge.

Another issue which needs to be understood when learning foreign language is vocabulary knowledge. Vocabulary knowledge refers to the knowledge of words as well as the meaning of the words. It requires the learners to use the words based on the context appropriately. The use of various words into good sentences and the selection of appropriate meanings based on context show that someone has learned foreign language. Nation (2001) describes the relationship between vocabulary knowledge and language use is that vocabulary knowledge enables language use and conversely the use of language refers to vocabulary enhancement.

Not only about word, but there is also a term called lexeme which will be found when learning English. Words seem to be the building-blocks of the language, because words can

stand alone and have meaning. Lexemes are not always same as words, words are units of language that can stand alone, but lexemes can represent several words and its inflection. Lexemes or lemma consists of a headword and its inflection, plural, third person singular, present tense, past tense, present participle, past participle, comparative and superlative, under the same part of speech (Nation, 2001).

In writing, the learners have to practice their vocabulary knowledge about words and lexemes based on what the writing tells about. In this case, the use of lexemes can be measured through lexical richness measurement. Kyle & Crossley (2016) stated that lexical richness is the measurement of how rich the students' have in writing and composing the words or lexical in a good writing. That statement proves that the lexical richness measurement applied in order to understand the way the learners use words and lexemes, and also how diverse lexemes are being used by learners. The result will be different depends on the understanding of each learner.

There are many kinds of writing works. One example of the writing works is short story. Short story is a type of literary works which tells about the story of character, all conflicts and the solution concisely. Short story tends to be dense and straightforward where there is a peak of the problem (climax) and its solution. Short story can be in the form of fiction or reality. Everyone can write all about experiences, hopes, imagination and others into a short story. During writing those stories, the learners of English language can be practicing their ability in writing.

English-language learners (ELL) are students who cannot communicate fluently, who usually come from non-English-speaking homes and backgrounds. They normally need specialized learning instructions in their academic courses. Because of Indonesia is one of non-English speaking country, the students in this country are considered as English foreign learners (EFL). Even the university students majoring in English are also as English foreign learners because they do not use this language in general society. They are expected to use English language both in speaking and writing. One example is the students of English Study Program in University of Trunojoyo Madura.

Writing is more essential in EFL students at university because they have enough knowledge to write and produce specific writing genres, such as non-academic writing. The students are required to write short stories with certain themes by using their own words. Since the English study program students of University of Trunojoyo Madura are EFL learners, then those short stories are considered as English learners' writing products. Those short stories are compiled into an anthology book of short stories. However, the female authors are more than the male authors. It can indicate that author's gender has a role to the writing.

This study focuses on the lexical richness of English learners' writings. Lexical richness in English is considered important because English is most widely used and studied as a main foreign language in many countries. Lexical richness shows the diversity of lexemes used by learners. In addition, this study examines lexical richness of the short stories according to the authors' gender. Singh (2001) stated that only a few studies have partially addressed gender differences in spoken and written language output. From that statement, this study initiates to add gender difference in this study. This study examines the anthology of short stories written by students of English Study Program, because that book consists of short stories which are likely to have the same number of words in every story and there are two genders who write those short stories.

Review of Related Literature

Lexical Richness

Lexical richness can be interpreted simply as the number of diction used by the author in an oral or written work. Diction means the choice of words that are appropriate (in their use) to express ideas so that a certain effects are obtained (as expected). The number of those words is counted through lexical richness measures. Lexical richness measures are possible to say that measure of lexical richness focus on how many different words are used in a text (Šišková, 2012). It can be determined by counting the different types in a text. Lexical richness measure has been developed in applied linguistics research. Many varieties of measuring techniques that can be applied in either written or spoken language have been proposed. Lexical density is measuring the amount of content words compared to the total amount of words in a text. Then, there is an assumption that the more difficult a word the less frequent it will be. Thus, a measuring of words appearance frequency in a text is known as lexical sophistication. The most frequently used for measuring the lexical richness is the type-token ratio (TTR). It is based on the different words (*types*) over the total number of words (*tokens*) in a text. This kind of measuring is known as lexical variation.

Lexical Density (LD)

Lexical density is one of the measurement techniques for analyzing the lexical richness in a text. Lexical density refers to the comparison between lexical words and the amount of words in a text. “Lexical density provides a measure of the proportion of lexical items (i.e nouns, verbs, adjectives, and some adverbs) in the text” (Johansson, 2008). A text is considered dense if it contains many lexical words relative to the number of words, since lexical words can convey the information of the text. Johansson (2008) also said that by investigating the research of lexical density, it can be received a notion of *information packaging* which means that a text with a high proportion of content words contain more information than a text with a high proportion of function words (prepositions, interjections, pronouns, conjunctions, and count words). The term of lexical words are more commonly known as content words or information words. “Content words are any nouns, verbs, adverbs, and adjectives which have suitable and significant lexical meaning” (Solichatun, 2011). It means that lexical density measures the information in a text according to the how tightly lexical words or content words have been inserted to the passage. It also provides the measurement of information in a particular piece of writing.

Laufer and Nation (1995) proposed the following formula to compute the lexical density of a text:

$$\text{Lexical Density} = \frac{\text{Number of lexical tokens} \times 100}{\text{Total number of tokens}}$$

The result of the calculation shows the density of the text that has been calculated. Solichatun (2011) stated that a high lexical density measures is around 60-70%, a quite lexical density measures is around 50-60%, and a lower lexical density measures is around 40-50%. The text that has percentage more than 60% is considered denser and also provides more informative words than the text that has percentage lower than 60%.

Lexical density has correlation with lexical diversity or lexical variation. However, the correlation may be inversely proportional. A text may have high lexical diversity but low lexical density. It is because of that text consist many of pronouns or auxiliaries rather than lexical items. This assumption is supported by statement form Johansson (2008) who states that a text may have high lexical diversity (i.e. contain many different word types), but low lexical density (i.e. contain many pronouns and auxiliaries rather than nouns and lexical verbs) or vice versa.

Lexical Sophistication (LS)

Lexical sophistication is also known as lexical rareness which measures the amount of advanced tokens in a text. Some words can be said as advanced words from the frequency of appearance in a text. The less frequent some words appear means that those words are advanced words. Advanced words tend to less used because considered as difficult words or not all people will easily understand about those words.

Lexical sophistication is also as one of the measurement techniques for analyzing the lexical richness. Read (2000) stated that it is kind of measurement for measuring the number of relatively unusual words or advanced words in the learner's writing. It is because the more difficult a word the less frequent it will be which is automatically influence the lexical richness.

In recent studies, Lexical sophistication is associated to the use of academic words in a text. Academic word is the word that commonly occurs in academic text. The Academic Word List (AWL) is developed by Coxhead (2000). AWL contains of 570 word families which are divided into 10 sublists. Each sublist has 60 words families, except for Sublist 10 which only has 30 words families. The sublists are sorted from 1 to 10 according to the word family's occurrence frequency. Sublist 1 consists of words families that appear most often, until Sublist 10 consists of words that rarest appear in academic text. Because of that reason, academic words are often used as a reference for measuring how sophisticated a text is.

Laufer and Nation (1995) proposed the following formula to compute the lexical sophistication of a text:

$$\text{Lexical Sophistication} = \frac{\text{Number of advanced tokens} \times 100}{\text{Total number of tokens}}$$

Sari (2019) concluded in her study that if a text has 9% or more percentage of Academic Word List (AWL), that text is considered as academic. This conclusion she got from many studies that showed the percentage of the AWL which are different from study to study. In other hand, Coxhead (2000) who has developed AWL stated that "the AWL accounts for approximately 10% of the total words in academic texts but only 1.4% of the total words in a fiction collection". It indicates that, for academic text the percentage of AWL should be 10% and for non-academic text the percentage of AWL is only 1.4%, in order to be considered to have sophisticated lexical.

Lexical Variation (LV)

Lexical variation is also known as lexical diversity or lexical range. It refers to the range of a learners' vocabulary as displayed in his or her language use. The measurement of lexical variation is the most used. Lexical variation is known as the type-token ratio (TTR). This measurement technique compares the different words or types over the amount of words in a text. Lexical variation can be determined simply by counting the different types in a text,

but in this case, it is clear that the number of types depends on the texts length and the longer the text, the more types it usually contains. (Šišková, 2012, p.28).

The statement above shows that, it will be difficult to compare the texts of different lengths. It may get different result. This measurement has been shown to be unstable.

Laufer and Nation (1995) proposed the following formula to compute the lexical variation of a text:

$$\text{Lexical Variation} = \frac{\text{Number of types} \times 100}{\text{Total number of tokens}}$$

Gender and Writing

Gender refers to the roles and responsibilities of men and women that are created in families, societies and cultures. Gender is the socially-constructed for men and women role. In social life, both men and women are inseparable from the use of language. Some social classifications such as ethnicity, gender, age, social class, and education determine the use of language by people. Gender is a concept that allocates different language skills to males and females according to context.

One of language skills is writing. Writing is interpersonal communication by using various styles of language. It is really related to everyone's personal life. By writing also allows everyone to communicate without thinking about time and distance. Moreover, writing skill is important in almost every course.

The study about gender in language skills is triggered from the previous articles. The pioneering work argues that women have a different way of speaking from men, because women speech typically displays some features such as hedges, super polite forms, tag questions, question intonation in declarative statements, empty adjectives, special lexicon, etc. (Alami et al, 2013). So that other studies were conducted to check whether women produce the same result in writing as what they do in speaking.

Method

This study dealt with the number of words that were used in the short stories and measurements, the design of this study was quantitative design. The source of the data is short stories written by students of English Study Program in 5th semester of academic year 2019, University of Trunojoyo Madura. The data were certainly the number of words (whole number) that have been taken from short stories by dividing them into some categories. Those categories were tokens, types, content words, and academic words list.

There are three applications that were used in this study, VocabProfiler to obtain the number of tokens and types for lexical text analysis, Claws Tagger to obtain content words by tagging the part of speech of each, and Academic Word List Highlighter to obtain academic words by highlighting the word. Lexical density result was obtained by dividing lexical words to the total number of words (tokens). Lexical sophistication was obtained by dividing the number of advanced words to the total amount of words (tokens). And lexical variation was obtained by dividing the number of word types to the total number of words (tokens). From the result of measurements, then it was determined the highest values, lowest value, and the average of each LD, LS, and LV.

Results and Discussion

Results

Table 1 shows the number of the tokens, types, content words and academic words of each short story. Those numbers of words were obtained by using some web applications. It consists of five columns. The first column contains of the list of the short stories' title. There are 48 short stories and they are sorted based on their original order in the anthology book. The second column shows the number of tokens in each short story. Tokens are simply to count every word which is spoken or written text and if the same word form occurs more than once, then each occurrence is counted. The third column shows the number of types in each short story. The fourth column shows the number of lexical words in every short story. The fifth column or the last column shows the number of academic words.

Table 1. Sample of Type, Token, Content Word, and Academic Word List

Short Story's Title	Token	Type	Content Word	AWL
The Miracle of Broken Plate	1078	381	666	9
George and The Origin of Victoria Village	1225	414	685	5
Love is Not Over	813	331	498	9
The Belle of Ruqayyah	814	301	476	6
Bloody Love: The Origin of Paterongan Village	953	347	559	14

Table 2 shows the classification of short stories based to the author's gender. For male authors, there are 14 short stories and for female authors, there are 34 short stories. There is a considerable difference from the amount of short stories that written by male and female, that is 20 short stories. This classification corresponds to the original source of data. There is no either addition or subtraction in the source of data. The writer only processed the data according to what is needed.

Table 2. Classification of Short Stories Based on the Authors' Gender

Gender	Amount	Short Stories' Title
Male	14	Bloody Love: The Origin of Paterongan Village; The Origin of Socah Village, Bangkalan, Madura; Sadikem of Jiwan Bridge; The Returnee of The Identity; The Legend of Surabaya; The Legend of The Sarangan Lake Magetan; Birthday Gift For Children; The Giant; Syekh Zainal Abidin (Sunan Cendana); Civil War; Mr. Hooman with His Miracle; The Origin of the City of Mojokerto, East Java; The Rebellion of Madurese Fighter Wizzard; and Srigati.
Female	34	The Miracle of Broken Plate; George and The Origin of Victoria Village; Love is Not Over; The Belle of Ruqayyah; You Are My Destiny; My Biggest Plight; The Piece of Heart; I Was Frightened All These Time; Behind My City; A

Sacred Place; The Magic of Lele; The Legend of Sacred Animals in Ngerong Cave; White Meliwis Bird Manifests Prabu Anglingdharma; Joko Sambang; Elanasewandono's Art Creation; The Legend of Kiai Raba; Lemah - Doro (Madura); Cenneng Stone; A Mysterious Woman in Pedeng; Rainbow Lake; Exorcist Dance; Legend of Lake Tondano in North Sulawesi; Chris Pain and The Giants; The Origin of The Name Bedilan Village; The Origin of Mengare Island; Nyai Agung Tumengkang Sari, Roro Jonggrang from Gresik; Nglirip Waterfall; The Origin of Sedudo Waterfall; Whiny Little Gitarja; The Legend of Banyuwangi; Mystery of Magic Fire; The Right of Left Book; Call Pak Jung in Kolpajung; and North Sumatra Folklore: The Wing.

Discussion

According to Laufer and Nation (1995), the most popular measures used in the description of the productive lexicon are lexical originality (LO), lexical density (LD), lexical sophistication (LS), and lexical variation (LV). However, in this measuring part, this study adopted the step used by Astridya (2018) that used three lexical richness measurements formula. Those measurements are lexical density, lexical sophistication and lexical variation in the short stories.

Lexical Richness

Lexical richness is simply interpreted to the use of the different words in a text. The most used for measuring the lexical richness is the type-token ratio (TTR). This measurement based on the use of different words over the amount of the total words (tokens) in a text. Lexical richness measurement begins to develop in applied linguistics research. Some measurement such as lexical density, lexical sophistication, lexical variation and lexical originality can be applied for measuring the lexical richness either in spoken or written language.

The previous part is about measuring and displaying the results of lexical density, lexical sophistication and lexical variation in the short stories. Those measures are obtained by using formula that proposed by Laufer and Nation (1995). After measuring and displaying the data, those results are analyzed in its relation to the lexical richness.

This study uses grouping system to obtain lexical richness result based on three previous measurements. The 48 short stories are grouped into 8 groups according to their each result in each measurement.

The first group consists of five short stories. Those short stories are *The Giant*; *Mr. Hooman with His Miracle*; *The Legend of The Sarangan Lake Magetan*; *Nyai Agung Tumengkang Sari*, *Roro Jonggrang from Gresik*; and *Behind My City*. They are grouped because of having high result in all three measurements that are lexical density, lexical sophistication and lexical variation. They have lexical density values of 61.31%, 60.03%,

60.54%, 60.16%, and 60.16% respectively. The average of lexical density is 60.44%. This result is categorized as a high lexical density because it has an average value more than 50-60%. Those five short stories have lexical sophistication values of 3.07%, 2.91%, 2.69%, 1.86%, and 2.47% respectively. The average of lexical sophistication is 2.60%. This result is categorized as a high lexical sophistication because the percentage of AWL is more than 1.4% of the tokens. They also have lexical variation values of 43.94%, 42.72%, 42.30%, 37.32%, and 36.59% respectively. The average of lexical variation is 40.57%. This result is categorized as a high lexical variation because it exceeds the average of overall lexical variation that is 36.40%. Therefore, those five short stories above are considered to have high lexical richness because they have high result in all three measurements.

The second group consists of three short stories. Those short stories are *Civil War; Mystery of Magic Fire; and Joko Sambang*. They are grouped because of having high result in two measurements that are lexical density, and lexical sophistication. They have lexical density values of 67.07%, 61.10%, and 60.99% respectively. The average of lexical density is 63.05%. This result is categorized as a high lexical density because it has an average value more than 50-60%. They have lexical sophistication values of 3.10%, 1.91%, and 2.30% respectively. The average of lexical sophistication is 2.44%. This result is categorized as high lexical sophistication because the percentage of AWL is more than 1.4% of tokens. However, their lexical variation cannot be said to be high. Their lexical variation values are 34.68%, 36.22%, and 32.98% respectively and make them have the average of lexical variation value of 34.63%. This result is not more than 36.40% or does not reach the average of overall lexical variations. Thus, those three short stories above have high result in two out of the three measurements that make them are considered to have a quite high lexical richness.

The third group consists of only one short story. This short story is *Love is not Over*. It has high values in two measurements that are lexical density and lexical variation. It has lexical density value of 61.25%. This result is categorized as high lexical density because it has a value more than 50-60%. This short story has lexical variation value of 40.71%. This result is more than the average of overall lexical variations that is 36.40%. However, its lexical sophistication value is 1.11%. This result does not reach 1.4%, the percentage of AWL in non-academic texts. So, this is not categorized as a high lexical sophistication. Therefore, the short story above has high result in two out of the three measurements that make it is considered to have a quite high lexical richness.

The fourth group consists of ten short stories. Those short stories are *Legend of Lake Tondano in North Sulawesi; The Origin of Mengare Island; The Origin of The Name Bedilan Village; The Origin of the City of Mojokerto, East Java; A Mysterious Woman in Pedeng; Elanasewandono's Art Creation; Bloody Love: The Origin of Peterongan Village; Birthday Gift for Children; Nglirip Waterfall; and A Sacred Place*. They are grouped because of having high result in two measurements that are lexical sophistication and lexical variation. They have lexical sophistication values of 3.04%, 3.45%, 3.02%, 5.29%, 1.79%, 2.92%, 1.47%, 1.55%, 2.60%, and 1.90% respectively. The average of lexical sophistication is 2.70%. This result is categorized as high lexical sophistication because the percentage of AWL is more than 1.4% of tokens. Those ten short stories have lexical variation values of 45.10%, 39.86%, 38.69%, 39.60%, 37.43%, 37.27%, 36.41%, 38.04%, 37.45%, and 37.23% respectively. The average of lexical variation is 38.71%. This result is more than the average of overall values of lexical variation that is 36.40%. Their lexical density values are 56.93%, 58.27%, 59.30%, 55.84%, 57.69%, 56.39%, 58.66%, 56.19%, 55.53%, and 54.31% respectively and make them have the average of lexical density value of 56.91%. This result is lower than 50-60% and not categorized as high lexical density. Thus, those ten short stories

above have high result in two out of the three measurements that make them are considered to have a quite high lexical richness.

The fifth group consists of two short stories. Those short stories are *The Miracle of Broken Plate* and *The Origin of Socah Village, Bangkalan, Madura*. They are grouped because of having high result in one measurement that is lexical density. The result shows that those short stories have high average lexical density, but in lexical sophistication and lexical variation are not categorized as high value. They have lexical density values of 61.78% and 60.95% respectively. The average of lexical density is 61.37%. This is categorized as high lexical density because it exceeds 50-60%. The result of lexical sophistication and lexical variation are quite low. Their lexical sophistication values are 0.83% and 0.76%. The average of lexical sophistication is 0.80%. The percentage of AWL does not reach 1.4% of tokens. It cannot be categorized as high lexical sophistication. Their lexical variation values are 35.34% and 30.48%. The average of lexical variation is 32.91%. This result is not more than 36.40% or does not reach the average of overall lexical variations. Therefore, those two short stories above have high result in only one out of the three measurements that make them are considered to have a quite low lexical richness.

The sixth group consists of nine short stories. Those short stories are *The Legend of Kiai Raba; The Piece of Heart; Syeikh Zainal Abidin (Sunan Cendana); Cenneng Stone; The Legend of Banyuwangi; The Returnee of The Identity; Exorcist Dance; The Rebellion of Madurese Fighter Wizzard; and Chris Pain and The Giants*. They are grouped because they have high result in one measurement that is lexical sophistication. The result shows that those short stories have high average lexical sophistication, but in lexical density and lexical variation are not categorized as high value. Those short stories have lexical sophistication values of 1.87%, 1.92%, 3.17%, 1.50%, 1.44%, 2.40%, 1.74%, 1.89%, and 1.66% respectively. The average of lexical sophistication is 1.95%. This is categorized as high because the percentage of AWL is more than 1.4% of tokens. The result of lexical density and lexical variation are quite low. Their lexical density values are 58.59%, 57.95%, 55.19%, 59.00%, 56.13%, 56.69%, 56.02%, 53.87%, and 56.37% respectively. The average of lexical density is 56.65%. This is lower than 50-60% and not categorized as high lexical density. Likewise, their lexical variation values are 35.02%, 35.33%, 34.86%, 30.83%, 32.76%, 30.35%, 30.93%, 28.87%, and 25.51% respectively. The average of lexical variation is 31.61%. It does not reach the average of overall lexical variations that is 36.40% and this is not categorized as high lexical variation. Thus, those nine short stories above have high result in only one out of the three measurements that make them are considered to have a quite low lexical richness.

The seventh group consists of nine short stories. Those short stories are *Sadikem of Jiwan Bridge; Call Pak Jung in Kolpajung; Lemah - Doro (Madura); The Legend of Sacred Animals in Ngerong Cave; The Right of Left Book; White Meliwis Bird Manifests Prabu Anglingdharma; The Belle of Ruqayyah; Whiny Little Gitarja; and I Was Frightened All These Time*. They are grouped because they have high result in one measurement that is lexical variation. The result shows that those short stories have high average lexical variation, but in lexical density and lexical sophistication are not categorized as high value. They have lexical variation values of 45.66%, 42.40%, 40.52%, 38.17%, 38.64%, 37.02%, 36.98%, 36.49%, and 40.10% respectively. The average of lexical variation is 39.55%. This is categorized as high lexical variation because the value is more than 36.40%, the average of overall lexical variations. The result of lexical density and lexical sophistication are quite low. Their lexical density values are 56.76%, 59.26%, 57.20%, 58.96%, 57.51%, 58.13%, 58.48%, 56.54%, and 51.78% respectively. The average of lexical density is 57.18%. This is lower

than 50-60% and not categorized as high lexical density. Then, their lexical sophistication values are 1.00%, 0.36%, 0.65%, 0.85%, 0.69%, 1.27%, 0.74%, 0.86%, and 0.71% respectively. The average of lexical sophistication is 0.79%. The percentage of AWL does not reach 1.4% of tokens, it cannot be categorized as high lexical sophistication. Therefore, those nine short stories have high result in only one out of three measurements that make them are considered to have a quite low lexical richness.

The eighth group or the last group consists of nine short stories. Those short stories are *North Sumatra Folklore: The Wing; My Biggest Plight; The Legend of Surabaya; Rainbow Lake; Srigati; George and The Origin of Victoria Village; You Are My Destiny; The Origin of Sedudo Waterfall; and The Magic of Lele*. They are grouped because they do not have high result in all measurements. The values of lexical density are 58.75%, 56.11%, 58.77%, 56.96%, 56.20%, 55.92%, 53.19%, 52.65%, and 54.24% respectively. The average of lexical density is 55.87%. This result is around 50-60% and categorized as quite high lexical density. Their lexical sophistication values are 0.98%, 1.36%, 0.66%, 0.43%, 1.00%, 0.41%, 0.30%, 1.09%, and 0.71% respectively. The average of lexical sophistication is 0.77%. This is said as low lexical sophistication because the percentage of AWL is no more than 1.4% of tokens. Furthermore, their lexical variation values are 35.74%, 35.82%, 32.75%, 34.43%, 34.33%, 33.80%, 35.07%, 34.01%, and 30.37% respectively. The average of lexical variation is 34.04%. This is also does not reach the average of overall lexical variations that is 36.40%. Thus, from explanation above, the nine short stories are categorized to have low lexical richness because all of their averages do not reach the standard for having high lexical richness.

Each short story has different result in the three measurements. It makes them are grouped based on the result. From the explanation of each group above, it can be summarized that there are four categories of lexical richness in this study. They are high, quite high, quite low and low. All of those are obtained based on the combination of three measurements that are lexical density, lexical sophistication and lexical variation. The high lexical richness category is for the short stories that have high results in three measurements. The quite high category is for the short stories that have high results in two measurements. The quite low category is for the short stories that have high results in only one measurement, and the low category is for the short stories that do not have high result in all measurements.

The majority of short stories that were written by English Study Program students of University of Trunojoyo Madura have a quite low lexical richness. This is proven by the existence of 20 short stories that included in this category with the percentage 41.67% from the whole. This is supported by only one lexical richness measure that has high value. Then, it is followed by the quite high category which occupies the second position. There are 14 short stories that included in this category with the percentage 29.17% from the whole. This is supported by two lexical richness measures that have high value. The third position is occupied by low lexical richness category. It is filled by 9 short stories with the percentage 18.75% from the whole. The last is high category. There are 5 short stories that are included in this category and fulfill 10.42% from the whole.

In this study, there are only a few short stories (i.e. 5 short stories) that have high lexical richness and the majority has quite low lexical richness (i.e. 20 short stories). This can be caused by sources of the data which are short stories. Unlike the previous studies, such as conducted by Astridya (2018) and Sari (2019) who investigated academic sources, this study investigates the non-academic text. The data sources of this study are short stories. Short stories are kind of non-academic text collection (fiction). Academic text uses academic language that has variety of words and more sophisticated vocabulary which are used for

specific purposes in a particular context. According to Coxhead (2000), word families of AWL are associated particularly with academic writing, so AWL occurs infrequently in fiction collection. Academic text also uses transition in sentences start such as ‘however’, ‘furthermore’, or ‘likewise’ instead of using conjunctions. Those can be the reason why in this study shows that majority of short stories has quite low lexical richness.

4.2.2 Lexical Richness based on Gender Differences

This section discusses the lexical richness based on gender differences which is between male and female authors of the short stories. This topic measures whether the amount of authors based on gender influence lexical richness. The words they produce might affect the lexical richness. There is a considerable difference from the amount of short stories that written by male and female. From 48 short stories, 14 of them are written by male and the remaining 34 are written by female. The short stories are divided into two group based on the author’s gender. The short stories that are written by male are *Bloody Love: The Origin of Paterongan Village; The Origin of Socah Village, Bangkalan, Madura; Sadikem of Jiwan Bridge; The Returnee of The Identity; The Legend of Surabaya; The Legend of The Sarangan Lake Magetan; Birthday Gift For Children; The Giant; Syeikh Zainal Abidin (Sunan Cendana); Civil War; Mr. Hooman with His Miracle; The Origin of the City of Mojokerto, East Java; The Rebellion of Madurese Fighter Wizzard; and Srigati*. Then, the remaining 34 short stories that are written by female are *The Miracle of Broken Plate; George and The Origin of Victoria Village; Love is Not Over; The Belle of Ruqayyah; You Are My Destiny; My Biggest Plight; The Piece of Heart; I Was Frightened All These Time; Behind My City; A Sacred Place; The Magic of Lele; The Legend of Sacred Animals in Ngerong Cave; White Meliwis Bird Manifests Prabu Anglingdharma; Joko Sambang; Elanasewandono's Art Creation; The Legend of Kiai Raba; Lemah - Doro (Madura); Cenneng Stone; A Mysterious Woman in Pedeng; Rainbow Lake; Exorcist Dance; Legend of Lake Tondano in North Sulawesi; Chris Pain and The Giants; The Origin of The Name Bedilan Village; The Origin of Mengare Island; Nyai Agung Tumengkang Sari, Roro Jonggrang from Gresik; Nglirip Waterfall; The Origin of Sedudo Waterfall; Whiny Little Gitarja; The Legend of Banyuwangi; Mystery of Magic Fire; The Right of Left Book; Call Pak Jung in Kolpajung; and North Sumatra Folklore: The Wing*.

The discussion of lexical richness is explained through the comparison of each measurement based on authors’ gender. The results of each measurement are presenting in the form of charts. The choosing of charts is related to the descriptive statistics which gives figures for all of the measures results. Hence, those charts are presented below.

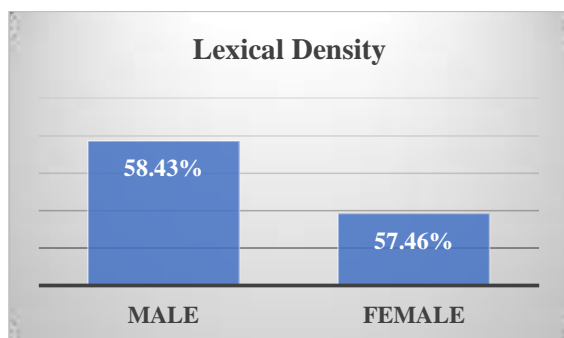


Figure 1. Lexical Density between Male and Female Authors

Figure 1 above shows the different result of lexical density between male and female authors. The male authors have higher result than female authors. From 14 short stories, male authors obtain average of lexical density values of 58.43% while the female authors obtain average lexical density 57.46% from 34 short stories. Both of those results are around 50-60%, so that they are categorized as high lexical density.

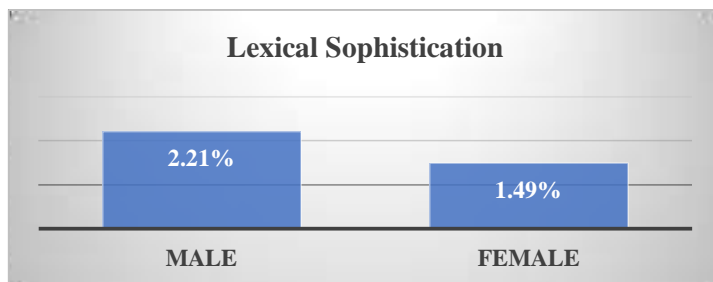


Figure 2. Lexical Sophistication between Male and Female Authors

Figure 2 above shows the result of lexical sophistication between male and female authors. In this lexical sophistication, male authors have higher result than female authors. The average of lexical sophistication results from 14 short stories written by male authors is values of 2.21%. In other hand, 34 short stories written by female authors have average of lexical sophistication values of 1.49%. For non-academic text such as short story, it is considered to have high lexical sophistication if the percentage of AWL exceeds 1.4%. Thus, both of genders are considered to have high lexical sophistication. Even though the female authors result is too close to the standard with the difference of only 0.09%.

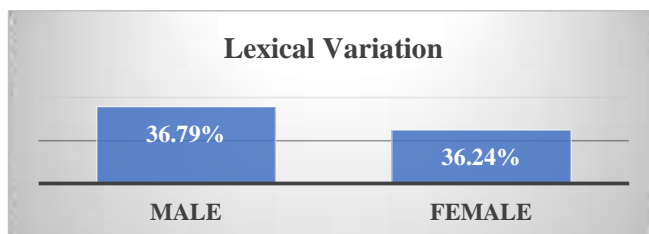


Figure 3. Lexical Variation between Male and Female Authors

Figure 3 above shows the result of lexical variation between male and female authors. As like lexical density and lexical sophistication results, this also shows that male authors have higher result than female authors. From 14 short stories written by male authors, it can be obtained the average of lexical variation values of 36.79%. Whereas, from 34 short stories written by female authors can be obtained the average of lexical variation values of 36.24%.

From the overall measurement results show that the lexical richness of short stories written by male is higher than those written by female. Even though the number of short stories written by female is more than male, the results show the opposite way. The results are similar to the study that conducted by Suganob-Nicolau and Sukanto (2016) that female authors produced longer sentences and more types than male in writing narrative essays, but this express the reality that the longer sentences are not directly related to the lexical richness. According to Ishikawa (2015), “men tend to use language to convey information or facts,

whereas women tend to use language to develop a rapport with other people.” This statement also supports the result of this study. The texts produced by male are not longer than female texts, but it has higher lexical richness. Male authors use more lexical words which convey the meaning and the information of what the texts tell, and also they use academic words for conveying the facts.

This part of discussion also clarifies that most of the short stories (i.e. 34 short stories) have lower lexical richness results, and the rest (i.e. 14 short stories) have higher lexical richness results. This is in line with the previous discussion which found that most of the short stories written by English Study Program students of University of Trunojoyo Madura have quite low lexical richness.

Conclusions

This study employs the theory and the lexical richness formula proposed by Laufer and Nation (1995) on lexical richness measurement in short stories written by students of English Study Program, University of Trunojoyo Madura. This study reveals three main points as a conclusion based on the discussion in chapter IV. First conclusion deals with the result of each measurement of lexical density, lexical sophistication and lexical variation. Second conclusion deals with the lexical richness of short stories from its lexical density, lexical sophistication, and lexical variation. Then the third conclusion deals with the lexical richness based on the author’s gender. This first point is the results of each measurement. There are 11 short stories that have high lexical density and 37 short stories that have lower lexical density. Then, there are 27 short stories that have high lexical sophistication and 21 short stories that have low lexical sophistication. Hence, there are 25 short stories that have high lexical variation and 23 short stories that have low lexical variation.

This second point is the results of lexical richness that obtained from the combination of lexical density, lexical sophistication, and lexical variation. There are four categories are formed in this study. They are high, quite high, quite low and low. The results show that quite low category occupies the first position which consists of 20 short stories. Then, quite high category occupies the second position which consists of 14 short stories. The third position is occupied by low category which consists of 9 short stories. The last position is high category which consists of 5 short stories. It indicates that the majority of short stories written by English Study Program, University of Trunojoyo Madura have quite low lexical richness.

This third point is the results of lexical richness of the short stories based on its authors’ gender. There are 34 short stories written by female and 14 short stories written by male authors. The results show that the lexical richness of short stories written by male is higher than those written by female authors. Because, in this study, the stories that written by male authors used more lexical words and academic words than the female authors. Female authors produce more words than male, but they do not use more lexical words and academic words. It means that most of the short stories are considered to have low lexical richness results. Even though the number of short stories written by female is more than male, the results show the opposite way.

Furthermore, it can be inferred from the conclusion points above that short stories written by students of English Study Program of University of Trunojoyo Madura have quite low lexical richness. Those results are not high enough. It is because data from this study are short stories which are considered as non-academic texts. The short stories (non-academic texts) are more imaginative and the language used is usually figurative or connotative,

whereas the academic texts have variety of words and more sophisticated vocabulary for specific purposes.

References

- Alami, M., Maryam, S., & Iranmanesh, M. (2013). Male-female discourse difference in terms of lexical density. *Research Journal of Applied Sciences, Engineering and Technology*, 5(23), 5363-5369.
- Astridya, F. W. (2018). Lexical richness of the expository writing in Indonesian by senior high school students. *Lingual*, 10(1), 23-29.
- Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, 34(2), 213-238.
- Crystal, D. (2003). *English as a Global Language* (2nd ed.). New York, USA: Cambridge University Press.
- Ishikawa, Y. (2015). Gender differences in vocabulary use in essay writing by university students. *Procedia*, 192, 593-600.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund University*.53, pp. 61-79. Dept. of Linguistics and Phonetics Working Papers.
- Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing* 34, 12-24.
- Laufer, B., & Nation, P. (1995). *Vocabulary Size and Use: Lexical Richness in L2 Written Production*. Oxford, UK: Oxford University Press.
- Moeller, A. K., & Catalano, T. (2015). Foreign Language Teaching and Learning. *International Encyclopedia for Social and Behavioral Sciences 2nd Edition*, 9, 327-332. doi:10.1016/B978-0-08-097086-8.92082-8
- Nation, I. S. (2001). *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press.
- Read, J. (2000). *Assessing Vocabulary*. Oxford, UK: Oxford University Press.
- Sari, W. A. (2019). *Lexical Richness and Syntactic Complexity of the English Exam Papers of SBMPTN*. Surabaya: Universitas Airlangga.
- Singh, S. (2001). A pilot study in gender differences in conversational speech on lexical richness measures. *Literary and Linguistic Computing*, 16(3), 251-264. doi:10.1093/lc/16.3.251
- Šišková, Z. (2012). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, 4, 26-36.
- Solichatun, S. (2011). *Content Analysis of Reading Materials in English on Sky Textbook for Junior High School*. Semarang: Walisongo State Institute for Islamic Studies.
- Suganob-Nicolau, M. F., & Sukanto, K. E. (2016). Gender differences in writing complex sentences: A case study of Indonesian EFL students. *Indonesian Journal of English Language Teaching*, 11(1), 69-80.