Corpus-Assisted Analysis of Robert Frost's Poem, "Into My Own" Using AntConc

*Amira H. Saad
ameera.saad@yahoo.com

Monaliza Sarbini-Zin
szmonaliza@unimas.my

Faculty of Language and Communication, UNIMAS
Sarawak, Kuching, Malaysia

ABSTRACT

This study aims to analyse the lexical features of one of Robert Frost's poems, 'Into My Own' using Laurence Anthony's AntConc. Based on a qualitative approach, the descriptive method was used to examine the information retrieved from the quantitative analysis of the literary text. The keyword method is to find out the most frequent words. Results show that corpus-assisted quantification of linguistic phenomena reveal more evidence than past studies which were intuitively analysed. Therefore, the findings indicate that using computer toolkits to analyse literary texts, mainly poems is possible by investigating lexical features to enhance themes, in this particular poem, the theme is a journey of independence. This study shows that the selection of only lexical features may seem insufficient but, eventually it is not possible to study all the poem's linguistic features.

Keywords: technology; corpus; Robert Frost; old methods; AntConc; linguistic features

ABSTRAK

Kajian ini bertujuan untuk menganalisis fitur-fitur leksikal salah satu puisi Robert Frost, 'Into My Own' menggunakan AntConc yang dibangunkan oleh Laurence Anthony. Berdasarkan pendekatan kualitatif, kaedah deskriptif digunakan untuk meneliti maklumat yang diperolehi daripada analisis kuantitatif teks kesusasteraan. Kaedah kata kunci adalah untuk meneliti dan menjana perkataan yang paling kerap digunakan. Dapatan menunjukkan bahawa kuantifikasi fenomena linguistik melalui analisis korpus mampu mendedahkan lebih banyak maklumat dibandingkan dengan analisis-analisis kajian terdahulu yang dilakukan secara intuitif. Oleh itu, dapatan juga menunjukkan bahawa penggunaan kit alatan komputer untuk menganalisis teks kesusasteraan, terutamanya puisi, boleh dilakukan dengan mengkaji fitur-fitur leksikal untuk pemurniaan tema-tema, di dalam puisi ini khususnya, temanya ialah pencarian diri sebenar. Kajian ini menunjukkan bahawa pemilihan fitur-fitur leksikal sahaja mungkin kelihatan tidak mencukupi tetapi, akhirnya, tidak mungkin untuk mengkaji semua fitur-fitur linguistik puisi tersebut.

Kata kunci: teknologi, corpus, Robert Frost, kaedah lama, AntConc, ciri linguistic

INTRODUCTION

Technology is developing and changing at such a rapid pace that it has become necessary for us as researchers and scholars to be constantly prepared to adjust and learn new methods to analyse literary texts. New developments in technology allow for more sophisticated softwares and more possibilities. Relatively few research studies have been conducted on poems (comparative to novels and drama), more specifically on corpus-based studies focusing on the linguistic features, namely, the lexical features which have been found to be worth examining. The present study is based on the assumption that the poem is rich in its lexis to represent the meaning of the poem, that is, the theme. Both nouns and verbs representing the lexicon of the poem, have multiple meanings, but, they are both often used to give the poem its broad and universal theme. Thus, the study aims to determine that lexical features in Robert Frost's poem, 'Into My Own' which help to convey the meaning. In fact, "Poetry is the result of the attitude of copyright poet born of the relationship between human life and human mind" (Omar, 2012). Robert Frost's poem, 'Into My Own' was selected in this study mainly because of the genius organization of the language, such as original lexical patterns. The original lexical patterns suitably present themselves to technological exploration. Specifically, in the present study, corpus linguistics-based method will be used to help the authors to have an examination of its linguistic features and interpret the poem's aesthetic value.

LITERATURE REVIEW

Robert Frost is an American-born poet who was celebrated for his portrayals of rural New England. Amid the American modern literary circle, Robert Frost is known as "American Poet Laureate" and is one of the greatest and the most popular poets. Based on the settings of New England, his poems that are full with pastoral scenery and his laboring experience are expressed through a comparatively traditional prosody. His unique writing technique and characteristic enable him to win more diversified audience since whether they are literary experts or demotic readers, they are capable to interpret the poems from dual perspectives of opaque modernism as well as a relatively more intelligible Victorian writing style. Frost's works from the early 20th century depicted rural life and its representations and explored issues of social and cultural diversity and philosophy. Robert Once wrote, "To the natural world, other people term my poem to art. However, I am not a poet for Nature; my poems involve much importance as well. Robert Frost utilizes Nature to demonstrate the internal battle people fight in their daily lives. His poetry typically starts by studying Nature and refers to the human condition, such as isolation, helplessness, uncertainty, and various social re- lationships. To Robert Frost's poetry, Nature is the most excellent characteristic of all. Frost has a profound passion for animals and an appreciation for design. He used to wander through the forest with his children and gazed before sleep into the starry heavens, from which Nature offered him a divine sense. However, Frost's poems do not dwell on tradi- tional rural life.

  Fagan (2007) describes 'Into My Own' as a poem that was composed early in Frost's life and has its roots in his personal life. In love with Elinor, Frost was a young man who was afraid to lose the love of his life, which led him to depression. 'Into My Own' was the opening poem of his first book, 'A Boy's Will' and was versed during that phase of his life in 1915. Walsh (1988) comments on 'Into My Own' as well as 'The Road Not Taken' that 'Into My Own' is ' one of only two poems he [Frost] wrote in England which has an avowedly English setting and subject' (p.211). Despite this fact, the poem has not been analysed stylistically enough. On the other hand, more attention has been paid to 'The Road Not Taken' (Ankrah and Nyamekye, 2017) and (Tyagi,2015), 'Stopping by the Woods on a Snowy Evening' (Widdowson, 2014) and 'Mending Wall' (Anggadewi,2020) as being his iconic poems. One point is worth mentioning here, in the perspective of the present time when the Covid-19 pandemic has massive impacts on our lives is and how the poets' memory can save our emotions about such pandemic that has struck the world out of a sudden. (Darek et.al,2021,p.3) that "contends that the poets have successfully expressed the impact of the COVID-19 pandemic on the life of the community whether in the terms of psychological, social, religious, cultural, educational, health and economic aspects". Comparatively, Robert Frost used the language of poetry to save in our memory his frustration

about the vague future of the world due to the woes of the World War I and his pessimism about what would happen to him and his ambitions in the future. Very often, the impact of catastrophes either it is a war, disease or a natural disease is never lost and saved by poet's words in their readers' memories forever.

Mahlberg (2007) notes that using computers to study literary texts is not new. However, "corpus approaches to literature and literary style are only in an early stage of development". According to Mahlberg, corpus stylistics investigates the relationship between meaning and form. Thus it is similar to both stylistics and corpus linguistics. Whereas stylistics pays more attention to deviations from linguistic norms that lead to the creation of artistic effects, corpus linguistics mainly focuses on repeated and typical uses, as these are what the computer can identify. (Mahlberg, 2007).

Corpus stylistics has made use of techniques to find the frequency and identify pattern to allow for in-depth analysis. Such techniques include the investigation of collocations, keywords and n-grams. Using this approach can help to make investigation between meaning and form more calculated and purposive. Robert Frost's poem, 'Into My Own' constitutes the data for this study. Technology has made it easier to retrieve information and analyse them thoroughly to reach to a more comprehensible theme.

The objectives of this study will be as follows:
- To generate a wordlist from the poem
- To find out the frequencies of the keywords and these relate to the theme of the poem
- To find word-concordance in this poem and demonstrate how this contributes to aesthetics

RESEARCH DESIGN

This is a corpus-based study, both quantitative and qualitive in nature, qualitative in terms of describing the retrieved information and quantitative because it applies statistics collected from the corpus toolkit.

CORPORA USED IN THIS STUDY

The corpora used are the text of 'Into My Own' and Brown Corpus.The Brown Corpus is an electronic collection of text samples of American English, the first major structured corpus of varied genres. This corpus first set the bar for the scientific study of the frequency and distribution of word categories in everyday language use. It is a general language corpus containing 500 samples of English, totalling roughly one million words, compiled from works published in the United States in 1961. It was a carefully compiled selection of current American English, totalling about a million words drawn from a wide variety of sources. Since Robert Frost is an American poet, his poems are appropriately compared with a corpus that includes American writings and works to display the characteristic features of his writing will be found.

The selection of a reference corpus is significant. It has an impact on all statistical procedures and outcomes. The size of the corpus and its substance are the most significant factors to consider while selecting a reference. The larger the reference corpus, the better and more precise keyword detection becomes (Scott, and Tribble, 2006). According to Scott (2010), a moderate reference corpus may be adequate; however, Sardinha (2004, quoted in Scott) and Tribble (2006) suggest that the reference corpus should be five times bigger than the node corpus, and that if it is larger, it will be more accurate. If the reference corpus is somewhat bigger than the node corpus, there is no need for keywords to be significant, and this is not a requirement. Here, the selected corpus is the Brown Corpus. Another reason why we preferred the Brown Corpus to the British Corpus is that the the poet(Robert Frost) in our analysis is an American origin.

On the other hand, "Words that are said to have a particular status, either because they communicate important evaluative social meanings or because they play a distinctive function in text or text - type," according to one of the study methodologies employed in corpus linguistics (Stubbs, 2010: 21). Keywords, according to this definition, are units that carry or contribute in uncovering the

meaning of a text from a linguistic standpoint; from a social one, they are units that have a partial reference to culture and society, and also symbolize the text type. There is a problem with the link between these two points of view.

RESEARCH PROCEDURE

The study goes through three steps as follows:

Step 1: Retrieve keywords and analyse the theme. The starting- point is the study of a text as opposed to the study of a whole language. Therefore, AntConc (3.5.9) tools are used, which enable the user to identify and study patterns in texts that are not easily found otherwise. The main tools are Wordlist, Keywords, and Concord. The Wordlist tool can generate word lists which are shown both in alphabetical and frequency order. The purpose of corpus toolkit is to locate and identify keywords in a given text. Any word that is found to be outstanding in its frequency in the text is considered "key". Concord is a program to be used to specify a search word. It will then present a concordance display, and give you access to information about collocation of the search word.

Step 2: Describe the linguistic features (in this study, the lexical features) systematically. Lexical features are data mainly extracted from the keywords database, and are categorised into themes, and abnormal usage groups.

Step 3: Demonstrate the aesthetic functions of the linguistic features. The ultimate aim of this research is to relate the objective linguistic data with reference to theme, and style.

FINDINGS AND DISCUSSIONS

'Into My Own' is Frost's extraordinary poem which explores our deep desire to break through the social and psychological boundaries in our lives which prevent us from achieving our ambitions and realising our 'true' selves. Frost's poem is in the form of a sonnet. Sonnets regained its popularity in the 20th. Century through the work of American poets, Robert Frost was among them. This structurally demonstrates the speaker's strict wish to retain his core self while reinforcing his beliefs.

LEXICAL FEATURES

The basic assumption is that there is a relation between word frequencies and the theme of the text. Keywords are divided into the following categories: the vocabulary related with the theme, the vocabulary showing the speaker psychology, words with abnormal usage.

Thus, AntConc is a good tool with a simple interface and as such probably best for people who are new to corpus linguistics. However, for the purpose of our present study, the researchers have preferred using AntConc. Having loaded our corpus into AntConc, it will be sensible to consider how to address our research question on wordlist and keywords and how to relate them to the poem theme and aesthetics, in general.

There is no single analytical route through which we can carry out corpus-assisted analysis, but we can employ a number of different procedures which can help us in our analysis. As has been already stated in the research objectives, we will employ the keyword analysis. A keyword is a word which occurs exceptionally frequently in a particular corpus. It is not necessarily one of the most frequent words in the corpus in itself, but its frequency is higher than we would expect if we compared our corpus against a second corpus. Such a comparison corpus often acts as some sort of reference, telling us what the typical frequencies of words are in the way a particular language is used. In order to find the keywords in our corpus we need to compare the word frequencies in our corpus against the frequencies in such a standard reference. For the study at hand, the researchers used the 6 million word written section of the Brown Corpus. When they (the researchers) run the keywords procedure, AntConc

obtains the frequencies of all of the words in both the corpus of the poem and the words in the written the Brown Corpus and then compares them together using a statistical test called a Log-Likelihood test. For each word in the "Into My Own " corpus a keyness score is obtained via the Log-Likelihood test, and the higher the keyness number the higher the likelihood that a word is a keyword. Thus, selecting a keyword could be based on its analysis which helps to answer our research question, and that its analysis reveals something that we did not already know or runs counter to expectations. Such keywords are likely to be of more relevance to our analysis.

Apparently, the fourteen lines disclose that some words have appeared more frequently than the rest of the words in the poem. Let us consider the poem by looking into the following statistics: the poem consists of 128 word tokens and 92 types. These words which have the highest frequency may guide us to overall meaning of the poem which can be revealed after the analysis. In other words, the microstructures (lexical elements) of a literary text generate its macrostructure (or overall meaning, i.e., the theme)
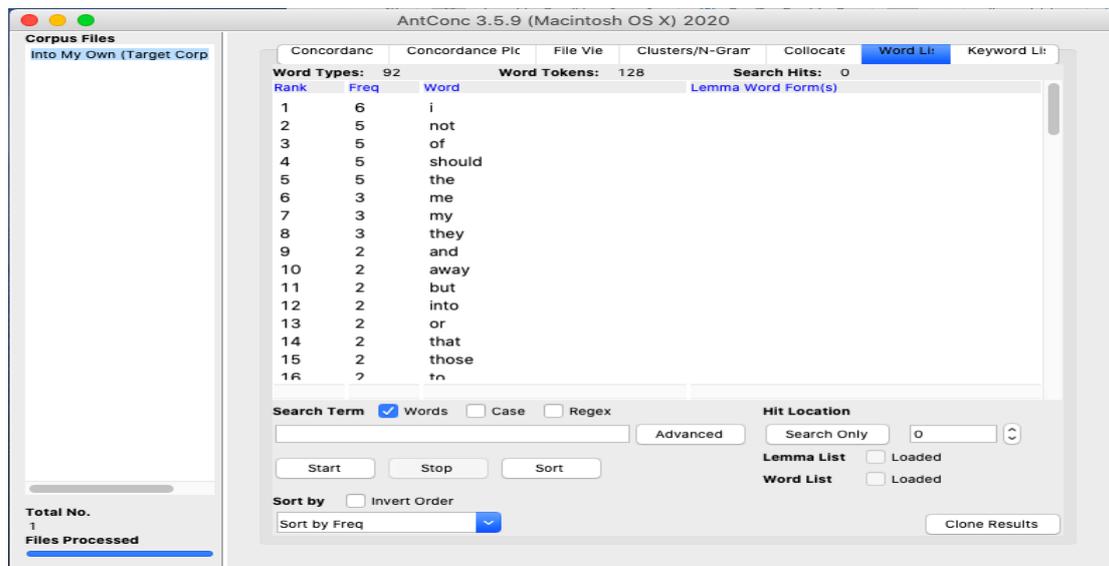

Figure 1. screenshot shows the word number, rank, frequency, and type

A keyword analysis is useful because it can reveal concepts in a corpus that are relatively frequent, and repeated usage is one way of signifying that a word is considered important to a text producer. We are often not able to accurately identify such keywords using qualitative techniques so they can provide a more objective way of helping us to narrow our research focus. The interesting part about this procedure is that it can be used on any corpus, even on a relatively small one in size. If our dataset is unfamiliar to us, then a keyword list can provide an overview of what some of the most salient concepts in the corpus are. If, on the other hand, we are already quite familiar with the texts in our corpus, the list might reveal some words that we had not considered as important. The concept of keyness is based on one of gradience – some words are more key than others but there is no definite cut-off point where we can say with certainty 'this word is a keyword but the next word on the list is not'.

The words in our corpus are ordered in terms of keyness and we then need to apply some sort of cut-off point in terms of deciding how many of them to consider for analysis. Practical considerations can delimit the number of keywords which are feasible to analyse, and they (the researchers) would advocate that a good analysis will cover a few keywords in-depth rather than produce a shallow, rushed analysis of lots of them and categorise these words into groups based on theme (as has been selected here)or grammatical class or semantic field. This will enable us to carry out a more in-depth analysis of a smaller number of keywords based on their connection to the theme. It is usually not necessary to spend a lot of time on two keywords that are used in similar ways (e.g. if the words red and crimson were keywords and were used essentially as synonyms in the same contexts. Other criteria for selecting a keyword could be that its analysis helps to answer our research question, and that its analysis reveals

something that we did not already know or runs counter to expectations. Such keywords are likely to be of more relevance to our analysis.

In terms of "Into My Own" corpus, the strongest 6 keywords are shown in Table 1, calculated for their Log-Likelihood keyness score using AntConc. All keywords had a keyness score of between 2767 and 7318, which indicate high keyness: a Log-Likelihood calculator gives a score of just 27.2 as being statistically significant at the 0.01% level[2] although if we had taken this score as the cut-off, it would have resulted in less than6 keywords. The numbers in brackets give the frequency of each word in the corpus and the researchers categorised the keywords into themes by hand. The following data are extracted from the word list above showing the frequencies of word occurrence:

Word	Frequency	Theme
I (First person pronoun)	6 times	Personae (poet)
Not	5 times	Negative aspects
Should	5 times	supposedly
of	5 times	belonging
they	3 times	unknown
me	3 times	The poet
my	3 times	Poet's possession

Table1. This table shows the highest frequency of the words in the poem and theme indicated

Since it is a soul journey to delve deep and search for his essence, the poet uses 'I' to show that the poem, 'Into My Own' can be relatable to all readers. The poet knows innately that this inner self is the true nature, and discovering that of himself will not change him into something different, but will only unveil who he really is. There are certain words used to indicate in a direct or indirect way the main theme of the text, these are called 'thematic signals'. These words have importance over others. They are simply the carriers of meaning or the clues to the hidden codes in the text. Roughly speaking, the poem's keywords have the following thematic events; loneliness, perplexity and fear of the future. The keywords extracted above have worked as the theme signals which enhance the theme in our minds. 'I' has been mentioned 6 times while 'me' and 'my'- both amounts six times. Apparently, Table 1 shows that words such as (I, my, me, it, etc) are dominant in terms of frequency. Nevertheless, this does not mean that researchers neglect the other words in the poem.

Computational retrieval of keyness should be checked and interpreted by manual examination to get a reasonable and a subtle way of reading the text by means of corpus stylistics study. While reflecting upon this, one can deduce that the fact that the poet is torn out between two realities, the first one is the circumstances imposed on him and made him helpless especially at that such early age (adolescence) and the second one is his possessions which he may leave behind or find behind the trees. Such perplexities transform his life into a difficult time. 'Not' sums up all the negativities collected in the poet's head and can be seen in words like 'doom' and 'gloom'. On the other hand, as it is guessed from the poem that the poet is confused between the actual reality and what is supposed to be in the future which can be all understood from the second high frequency of the word, 'should'. 'They' as opposed to 'I' and talks about the unknown part of the journey and also as half frequent as 'I'. It can be inferred from such statistical use that the possibility of facing the unknown usually bears 50% of the expectations which is the same percentage between 'I' and 'they'. The following table displays how the most frequent words after being compared with the Brown Corpus have provided the keyness and their effect.
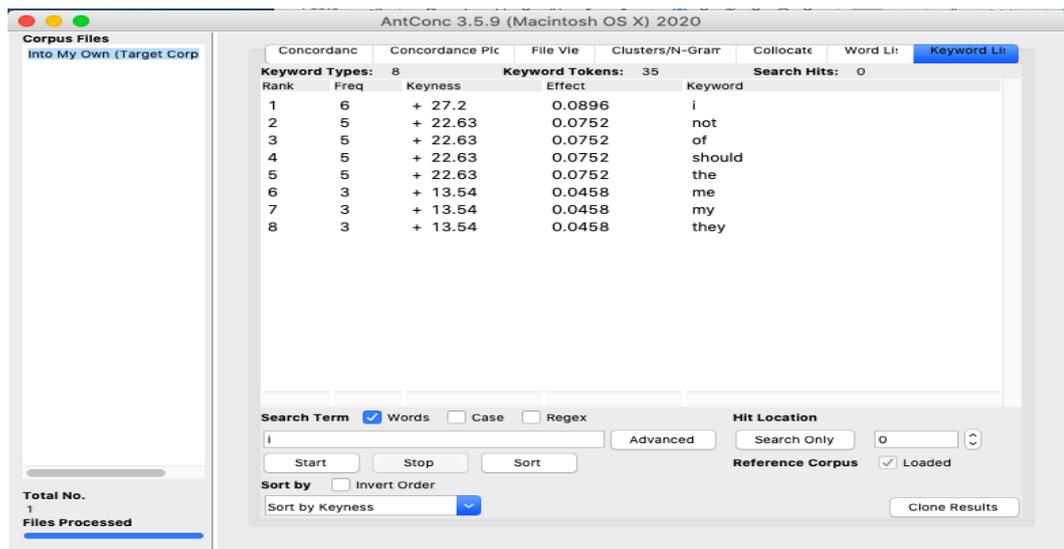
Figure 2. Screenshot Shows the keyness and the effect of the highest frequent words.

In general, keywords are computed using two word lists, one from the text or study corpus that one wants to investigate and the other from a normally larger, reference corpus that acts as a benchmark corpus or provides background data for keyword calculation. Comparing with the Brown Corpus, some words are shown to be favourites with Frost in his pessimism mirrored in 'Into My Own'. Adjectives are frequently of negative meaning, or used in negative context, such as (dark, old, firm, merest, slow, and long). Though the poem bears signs of youthful aspiration of an enthusiastic poet, it established pattern of ambiguity. Specific word choices are used with specific purpose. The reader slowly realises that Frost leaves more questions unanswered than answered; yet the poem gives a sense of satisfaction and fulfillment.

The poem is divided into two tenses the present tense and the unreal past which also indicates his confusion about what he wants and what is supposed to be and what he wishes to happen. While brave enough to steal away into the vast unknown, he is not totally sure of what he will find. This unfolds archetypical, mortal human nature, always groping in darkness: the darkness of innocence and ignorance as well.

CONCLUSION

This paper argues that the use of corpus in the analysis of a literary work is motivating and rewarding. The poem is an index to the choice of theme Robert Frost loved to write on. The poem airs the poet's inner-most fancy and deep feelings of his robust optimism and justifies its placing at the opening of his newly volume. Though, it is considered to be a less known poem, it does manifest the poet's firm conviction and mastery of stylistic devices. Commenting on the poem, Lawrence Thompson(1966) notes that Frost "gradually achieved perspective on these earlier lyrics which enabled him to view some of them with ironic ambivalence." All these details cannot be revealed or comprehended without the help of AntConc which shows how technology and aesthetics can work together to add to our knowledge about the previously studied literary texts. Corpus analytic techniques are multi-purpose strategies with immense potential to enhance all sorts of textual analyses and to confirm or contradict our intuitions about patterns and meanings in the literary and non-literary language.

In our corpus-stylistic analysis, we aimed to offer some revealing insights into the frequency of the keywords, the related key lexical items and word combinations employed in Robert Frost's poem. From the vantage provided by the computational corpus tools and analysis, we may conclude that this poem written by Robert Frost during the World War I is characterized by the following stylistic features: (1) an individual perception of life; (2) realistic pessimism; (3) a concise factual style marked by a use of textual ambiguity hidden in the language simplicity. Evidently, the results that we have drawn from

our corpus-stylistic analysis are by no means exhaustive enough to identify and explicate all the linguistic features in the poem, but we believe it may help provide a substantial critical ground for further corpus-assisted studies.

IMPLICATIONS AND LIMITATIONS

As no research is ever definitive or exhaustive, this one is still open. The selected features in the present analysis have been those of high frequency and salience to the writer, but one should be aware all the time that the observations have been subjective and therefore are bound to be incomplete, as some important features may not occur that often. In other words, there is still room for more detailed study, since no critics can exhaust a literary work's value. Moreover, the users of the AntConc can be divided into beginners and specialist. Depending on who uses the linguistic toolkit, the analysis may be different.

Accessing proper corpora and knowing how to make the most of corpus tools might be difficult for beginning researchers. When you have a variety of processes at your disposal, it might be difficult to decide which ones to utilize and in what order. Beginning analysts may rely on a tool's default settings or place too much emphasis on reporting frequencies that may or may not disclose much of interest, rather than doing more extensive concordance analysis. It's vital to remember that a keyword is merely the beginning, a 'gateway' into the corpus, and that it should always be followed up by qualitative studies. Furthermore, the analysis should ideally extend beyond the texts.

As a result, analysts may become overwhelmed by the diverse set of abilities necessary to conduct and combine the many quantitative and qualitative parts of the study. Furthermore, analysts who are not linguists may be tempted to do a content study of the corpus, which may fail to reveal how language is employed to preserve or dispute representations or discourses.

APPENDIX

Into My Own
by Robert Lee Frost

One of my wishes is that those dark trees,
So old and firm they scarcely show the breeze,
Were not, as 'twere, the merest mask of gloom,
But stretched away unto the edge of doom.
I should not be withheld but that some day
into their vastness I should steal away,
Fearless of ever finding open land,
or highway where the slow wheel pours the sand.
I do not see why I should e'er turn back,
Or those should not set forth upon my track
To overtake me, who should miss me here
And long to know if still I held them dear.
They would not find me changed from him the knew-- Only more sure of all I though was true.

REFERENCES

Anggadewi, K. N. (2020). Stylistic Analysis in Robert Frost's "Mending Wall*". Master Thesis.*Universitas Sanatha Dharma, Yogyakarta.

Ankrah, G. K., Kwasi, O., & Nyamekye, P. (2017). Lexical Choices in "The Road Not Taken", A Poem Written By Robert Frost (1874-1963). European Journal of English Language Teaching.

Anthony, L. 2020. *AntConc* (Version 3.3.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Derak, R., Manap, N. A., & Ahamad, A. (2021). Impak Pandemik COVID-19 dalam Penulisan Puisi di Malaysia. *PENDETA: Journal of Malay Language, Education and Literature*, *12*, 49-57. https://doi.org/10.37134/pendeta.vol12.edisikhas.5.2021

Fagan, D. J. (2007). Critical Companion to Robert Frost: A Literary Reference to His Life and Work. Infobase Publishing.Frost, R. (1969). Into My Own. The poetry of Robert Frost, 5.

Hart, C. (2020). Researching Discourse: A Student Guide. Routledge.

Mahlberg, M. (2010). Corpus linguistics and the study of nineteenth-century fiction. *Journal of Victorian Culture*, *15*(2), 292-298.

Omar, A. (2012). Nukilan Sajak Merdeka dalam Membangunkan Bahasa, Sastera dan Budaya Bangsa melalui Pengajian Kesusasteraan Melayu. *PENDETA: Journal of Malay Language, Education and Literature*, *3*, 195 - 213. Retrieved from https://ejournal.upsi.edu.my/index.php/PENDETA/article/view/1123

Scott, M. (2010) Problems in investigating keyness, or cleaning the undergrowth andmarking out trails, in M. Bondi, and M. Scott, (eds.) Keyness in texts, Amsterdam:John Benjamins.

Scott, M. and Tribble, C. (2006) Textual patterns: key words and corpus analysis in language education, Amsterdam: John Benjamins.

Thompson, Lawrence (19660. Robert Frost: The Early Years (1974-1915)

Tyagi, A. (2015). An Analysis of Robert Frost's Poem:"The Road Not Taken". PhD Scholar, SRM University, Sonepat." Veda's journal of English Language and Literature-JOELL, 2, 66-68.

Walsh, J. E. (1988). Into My Own: The English Years of Robert Frost, 1912-1915. Grove Pr.

Widdowson, H. G. (2014). *Stylistics and the Teaching of Literature*. Routledge.

Wu, H. (2021, October). Duality in Robert Frost's Poetry: The Fusion of Tradition and Modernism. In 2nd International Conference on Language, Communication and Culture Studies (ICLCCS 2021) (pp. 393-399). Atlantis Press.