Corpus-Based Study: Exploring The Frequency of 'Nouns' Used in Poetry Written by UiTM Undergraduates

Nur Baiti Ismail Shauki
baiti.is@umk.edu.my
Centre for Language Studies and Generic Development
Universiti Malaysia Kelantan

Manvender Kaur
manvender@uum.edu.my
School of Education and Modern Languages
Universiti Utara Malaysia

ABSTRACT

This study aims to investigate the frequency level of 'nouns' used in poems written by the undergraduates of Universiti Teknologi MARA Kampus Machang based on Paul Nation's headword list. The importance of knowing the frequency will determine teacher's priority in teaching vocabulary. The effect of teacher's implementation of standard nouns in classroom should reflect in the target behavior of the learner as they utilize the word in writing and mostly in communicating. The methodology used in this mini-scaled investigation of the written communication is the Computer-Assisted Corpus Analysis (CACA). Ten poems written by ten undergraduate students were collected and analysed by Tagger C5 tagset to identify the 'nouns' and separate it from other content words. Subsequently, 'nouns' frequencies were calculated by using AntConc software in order to determine the type of noun that most frequently occurred in the student's writing. The list of 107 words was then compared to Paul Nation's English Vocabulary Profile (EVP) program of academic vocabulary to determine the level of vocabulary that the students possessed. In the discussion section, the result of the frequency is analysed to determine whether the level is at par with the standard vocabulary knowledge or below. Although the concordance overall results show that 'nouns' was the most frequent used word in the poems,

Keywords: nouns; Computer-Assisted Corpus Analysis (CACA); written communication; part-of-speech (POS) tagging

INTRODUCTION

Generally, following Paul Nation's methodology, high frequency words are of 2,000 words and low frequency word is around 10,000 in number. Nation (2001, p. 388–389) mentioned "[i]f more than five percent of the running words are unknown, then it is likely that there is no longer meaning-focused learning because so much attention has to be given to language features". Undoubtedly it is of utmost necessity to investigate the frequency level of vocabulary use in language lesson to determine its relevance. If the frequency level is incompatible with the standard, it is necessary for the teachers to provide supplementary activities to bridge this gap. Language learners would be at loss without this kind of activities as their vocabulary level is insufficient to assist them in understanding materials of high frequency words.

Upholding the latest government's educational policy which is initiating the Common European Framework of Reference for Languages (CEFR) in teaching and learning English Language, this study compares the vocabulary that the students utilise in their writing to the list of CEFR based vocabulary in English Vocabulary Profile website. The English Vocabulary Profile (EVP) is a free online resource that chronicles the CEFR levels of words, phrases, phrasal verbs and idioms. Thus, CEFR and EVP share common aim to explicate the student's gradual development of learning English. In the Malaysian Educational Plan, students are expected to perform based on the CEFR scales as the benchmark. Yamat, Umar and Mahmoud (2014, p. 198) propose that CEFR is 'a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop to be able to act effectively'. While the Council of Europe explains further in the CEFR manual, "... a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe. It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively. The description also covers the cultural context in which language is set. The Framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis" (2001a, p. 1). For undergraduate students, B2 is the minimum level of requirement especially for admission to United Kingdom's prestigious university such as Cambridge. Thus, it is significant to explore what is the frequency of Nouns used by Malaysian undergraduates' in their writing.

This paper explores the frequency levels of 'nouns' used in undergraduates poems as nouns serve as the spine of sentence structure. Goldin-Meadow, Butcher, Mylander, Dodge (1994) confirm nouns role as "it also appears to be central to the structure and function of language" (p. 259). Other researchers also agree on the significant role plays by noun although it is hardly been separated from the verbs. For example Sapir (1921, p. 119) considers (nouns) to be essential to the life of language. Baker (2003) later on defines a noun is a lexical category that introduces a referential index. Therefore, this study utilises Corpus Linguistics to examine the explicit area of English grammar, namely, nouns. The purpose is to explore the means Malaysian undergraduate students use nouns in their poem writing. Writing poetry is the teacher's own initiative to make the class activity more fun and less stressful rather than writing a lengthy composition when teaching noun.

The university is chosen because it is in the setting that the author of this paper works. The research questions which direct the current study are as below:

(a) What is the frequency of the nouns used by undergraduate students in their poetry writing?
(b) What category of nouns is mainly used by the undergraduates in their poetry writing?
These questions will be answered in the discussion section.

The Importance of Frequency and Corpus

Paul Nation's English Vocabulary Profile is a convenient application for the matter of word rank. The list is organised in such a way that the level of the words and learners description is catalogued accordingly. Other than sequencing the standard of English of Malaysian undergraduate, frequency is another function of EVP that is treated as equally significant to be explored. In understanding L2 vocabulary instruction, it is necessary to select from the frequently use word, which words are of a priority. As Nation and Gu assert that "in terms of usefulness, all words are not created equal" (Nation, 2007, p. 20). The idea of 'usefulness' in this perspective refers to communicative usefulness. In this modern era, the main and vital goal of language learning is for the students to be able to convey themselves successfully. Communication is made better with the presence of appropriate vocabulary that is utilised within context. Parenthetically, vocabulary learning should expose the students with the words that bring greatest advantage in terms of expressing themselves sufficiently.

Among the advantages offered by knowing highest frequency words is it will inexorably linked to the communicative efficiency of the user. As Baron, Rayson and Archer (2009) states, "A frequency list records the number of times that each word occurs in the text. It can therefore provide interesting information about the words that appear (and do not appear) in a text". For instance, the word *smile, smiling, smiled and smiles,* do we count it as one word or six parental words? Thus, other that providing a data that stores regularity of the word, frequency list also contributes abundance of information of the implicit or explicit word. Moreover, it is essential to teach the learners knowledge on vocabulary that provides them with the utmost benefits. Nation (2001) states that major interest should be directed to the highly frequent words in General English. To be exact, low frequency words are not worthy of classroom time, and different learning and guessing strategies should be applied after sufficient exposure on those frequent words.

Gries (2009, p. 2) notes that frequencies are the based of corpus linguistics, thus it is significant to discuss the concept of corpus being the data provider for frequency analysis. The definitive concept of corpus as a "collection of written or spoken texts" (Wehmeier, 2000, p. 295) has been an accepted notion for quite some time until in the 20th century new meanings emerge in the area of language studies. Corpus, as indicated by Tognini Bonelli (2001, p. 2), is "a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis". Therefore, corpus in definition is data collected from various sizes of samples is compiled for the purpose of further investigation. Regarding its benefits to teaching and learning, McCarten (2007) lists 6 usage of corpus:

1. Frequency: Which words and expressions are most frequent and which are rare.
2. Differences in speaking and writing: Which vocabulary is more often spoken and which is more often written.
3. Contexts of use: The situations in which people use certain vocabulary.
4. Collocation: Which words are often used together.
5. Grammatical patterns: How words and grammar combine to form patterns.
6. Strategic use of vocabulary: Which words and expressions are used to organize and manage discourse.

These data provide linguists opportunity to make thorough and systematic studies of how language is actually utilise. Thus, the current research reflection on students' authentic writing should create resemblance of what level does the students vocabulary level is at. By relying on such data, teachers can further analyse their student's learning pattern, and discover, for example, what elements should be focused on when teaching the student a specific language

skills. Therefore, corpus-based studies are of great aid to language teachers. Once teachers are conscious of their students' ability, they will be able to guide them to accomplish teaching objectives.

METHODOLOGY

This chapter describes (1) the research design, (2) sampling, (3) data collection, (4) data analysis, and (5) validity and reliability. First, the research design section will define the type of research design, the population and sample, the instrument, and the procedures used for the study. Then, the data analysis section will define all the variables used in the study, as well as describe the statistical analysis process of the study. Finally, the validity section focuses on the reliability and validity of the instruments, as well as the research study as a whole.

Research Design

The research design of this study is descriptive mixed methods because the data of the research contains both numbers and words. Sandelowski, Voils, & Knafl (2009) once stated that qualitative results can be compared with the quantitative dataset. Thus, even though the presented data were in form of words, there were some numerical data obtained by calculating the total sum of personal nouns used.

Black (1994) once stated that qualitative methods take a holistic perspective which preserves the complexities of human behaviour. Thus, in this study, the writer used quantitative method at first, after collecting the data, the writer began to analyze the data qualitatively. In analyzing the data, the writer used the following steps:
1. read the frequency data.
2. comparing the frequency of nouns and its types
3. finding the used of nouns in the poems.
4. concluding the data.
5. reporting the result.

Sampling

The poems used were written by ten first year Diploma students of Universiti Teknologi MARA Kampus Machang. Age range is from 18-19 years old and all of them enrolled in Accountancy course. Their English proficiency level is intermediate or lower B1 if we go by CEFR rank. The reason of selection of only these 10 students' writing is due to the originality of their ideas and the rest of the writing contain errors that hamper understanding.

Data Collection

At Malaysian tertiary institutions in general, it is observed that many learners do not possess the basic word knowledge necessary for successful tertiary education (Kaur, 2007). Thus, this study attempts to explore Malaysian undergraduates' frequency level of vocabulary, whether or not they are at the required standard of lexical knowledge or below. In order to attain the result of their frequency level, the vocabulary used in their written assignments is compared to Nation's English Vocabulary Profile (EVP).

There were around 25 students in this class and all of them were in their first year in Diploma of Accountancy course. They took BEL 120 or Consolidating Language Skills as one of the course requirement for graduating. This skill is deliberated for learners whose proficiency in English language is at the beginner or transitional levels. It endeavours to elevate their level of proficiency in the language to the high intermediate level. This aim is

accomplished through the integration of the four skills of reading, writing, speaking and listening, with suitable concern placed to exposing an appropriate operational understanding of intermediate-level grammar. The credit hour is 3 and among the course content are part-of-speech, subject-verb agreement and verb tenses.

The poem writing exercise was supposed to enhance their understanding on the use of nouns taught in the two hours class. The first one hour was dedicated to equip them with knowledge on different types of noun, namely *abstract noun, proper noun, common noun, collective noun, countable and uncountable noun*. Moreover, the lesson also covered the topic of nominal quantity which deals with *singular, plural and neutral noun*. After the exposure is given on the topic, the students were given 1 hour to come up with a poem that portrays correct use of nouns reference. The researcher chose this class because the students had documented the writing assignment online by submitting it to the researcher's email. Only ten poems were chosen as these ten had shown correct use of nouns. Therefore, the researcher is interested to find out what is the frequency level of the nouns that they have used in their writing and is there certain types of nouns that frequently occur in their writing.

Data Analysis

CLAWS (the Constituent Likelihood Automatic Word-tagging System) tagger

After ensuring all students has submitted the exercises, the analysis of the data was carried out. Firstly, the writer retrieved the copy of the poems from the email. Each students has to submit both the soft copy and the hardcopy of their poem to the researcher immediately after they finish writing it. The researcher then used CLAWS tagger, a computer-based learner corpus analysis for POS tagging (Appendix B), to count noun frequency which tagged as *nn*. CLAWS tagger was constructed by the UCREL of Lancaster University in the early 1980s and it is available online for free. The system has an error rate of 1.5% and therefore has consistently achieved the accuracy of 96-97%. POS tagging is the process of marking a word in a text as corresponding to a particular part of speech based on its definition and context (Figure 1).

There are two sets of output available in this application, the smaller output is Tagset5 and the larger output is Tagset7 in which one can access more words of British National Corpus. This study is using tagset5 for the reason explained in Findings section later. The limit for text submission in tagger box is 100 000 words. Once the text has been typed or pasted in the form, choose 'Tag text now' below the box (Figure 2). As the writer intended to find out the grammatical items in the poems, CLAWS tagger provided the platform to do as such by listing out all the grammatical items contained in the poems. For this study, the focus will be on *nn0* or *neutral noun, nn1* or *singular noun, nn2* or *plural noun* and *np0* or *proper noun* (Figure 3).

AJ0	adjective (unmarked) (e.g. GOOD, OLD)
AJC	comparative adjective (e.g. BETTER, OLDER)
AJS	superlative adjective (e.g. BEST, OLDEST)
AT0	article (e.g. THE, A, AN)
AV0	adverb (unmarked) (e.g. OFTEN, WELL, LONGER, FURTHEST)
AVP	adverb particle (e.g. UP, OFF, OUT)
AVQ	wh-adverb (e.g. WHEN, HOW, WHY)
CJC	coordinating conjunction (e.g. AND, OR)
CJS	subordinating conjunction (e.g. ALTHOUGH, WHEN)
CJT	the conjunction THAT
CRD	cardinal numeral (e.g. 3, FIFTY-FIVE, 6609) (excl ONE)
DPS	possessive determiner form (e.g. YOUR, THEIR)
DT0	general determiner (e.g. THESE, SOME)
DTQ	wh-determiner (e.g. WHOSE, WHICH)
EX0	existential THERE
ITJ	interjection or other isolate (e.g. OH, YES, MHM)

FIGURE 1. The tagset5 of CLAWS Tagger application

Select tagset: ○ C5 ○ C7

Select output style: ○ Horizontal ○ Vertical ○ Pseudo-XML

Type (or paste) your text to be tagged into this box.

Tag text now   Reset form

FIGURE 2. The form for text submission

NN0	noun (neutral for number) (e.g. AIRCRAFT, DATA)
NN1	singular noun (e.g. PENCIL, GOOSE)
NN2	plural noun (e.g. PENCILS, GEESE)
NP0	proper noun (e.g. LONDON, MICHAEL, MARS)

FIGURE 3. Tag focus of the study

AntConc software

The next step is to collect the tagged data in CLAWS Tagger application. The data which is in Microsoft Word is then converted to plain text format. Then, by using AntConc application version 3.2.4w, a freeware corpus analysis toolkit that can be downloaded for free from the internet, the researcher uploaded the tagged poems into the application. AntConc is a software developed by Laurence Anthony of Waseda University. Among the many functions of this program is organising the words alphabetically and sorting the frequency from the files loaded.

It also provides reverse ordering, word endings ordering and case-sensitive word listing (Figure 4). The application then analysed the poems for the total number of each words used by the students for example, the word 'the' was used 27 times in all ten poems and the word 'to' is 18 times. Hence, the writer was able to obtain data of the frequency of nouns used in the poems by looking at the total number of *nn* tag (Figure 5).
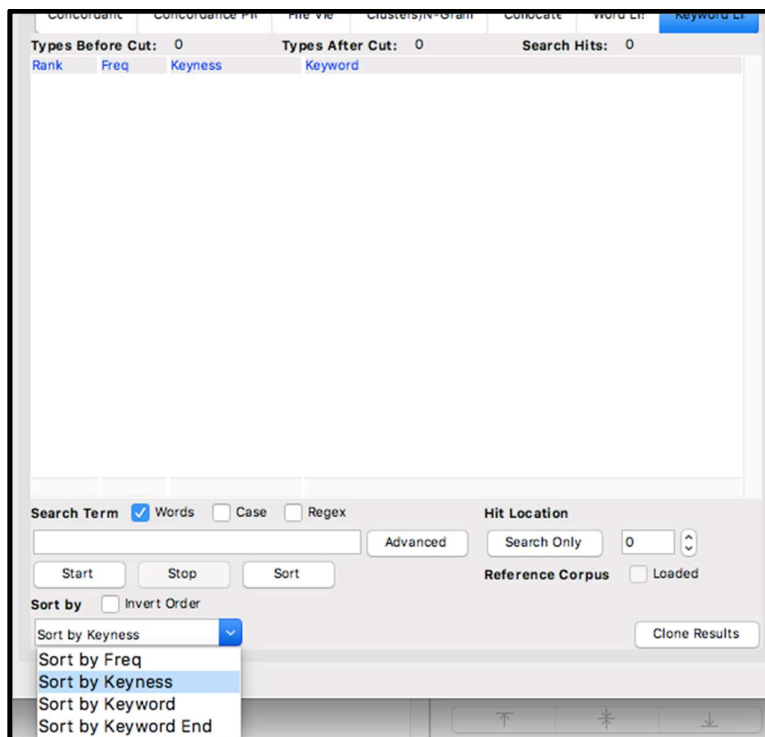

FIGURE 4. Keyword sorting

FIGURE 5. The frequency of *nn* code

English Vocabulary Profile

After analysing the data using CLAWS Tagger and AntConc software, the researcher then uses English Vocabulary Profile (EVP), an interactive web resource which provides a huge searchable database of detailed information on the words and phrases that learners know at each level of the Common European Framework (CEF).

The English Vocabulary Profile is a website that provides free access to the research findings on what English vocabulary and grammar is suitable for teaching according to each CEFR level. The list of words is based on extensive research using the Cambridge Learner Corpus (CLC), a multi-billion word corpus of spoken and written current English, covering British, American and other varieties. It collected data from learners all over the world particularly several hundred thousand examination scripts written by learners from all over the world that is added to every year. The research was led by two departments of the University

of Cambridge, UK: Cambridge University Press and Cambridge English Language Assessment. Information on the A1 - C2 levels is also provided for individual meanings of words, to give an indication of teaching and learning priorities.

As it is important to know the level of vocabulary that Malaysian undergraduate students possessed, EVP offers many functions that can be used by any researchers or educators depending on the part of language to be assessed. There is a British and American English version, and there is audio pronunciation for all entries. User can choose to what extent of CEFR level that they want the word to list; A1, A1 - A2, A1 - B1, A1 - B2, A1 - C1 and A1-C2. There is also an option for single level search of whether to view A2 only, B1 only, B2 only, C1 only or C2 only (Figure 6).

FIGURE 6. Single level search option

FIGURE 7. Advanced option

Moreover, by leaving the text box blank and selecting one of the levels, one can generate lists of words for each level. User can also filter the list using the advanced search functions, for instance to find out what 'homes and buildings' vocabulary A1 learners know, or which phrasal

verbs are known at B2 level. Advanced search option also enable user to use the wildcard (*) search to discover, for example, which '-ing' adjectives are known at A2 (see example in Figure 7).

Therefore, for this paper, after analysing the noun frequency, researcher list down the highly frequent nouns in student's poem and key-in the word in the search box. Prior to that, option A1-C2 level is chosen to accommodate the search of undergraduates levels of vocabulary.

Validity and Reliability

The methodology used to analyse the data has a very low rate of error therefore, its reliability is high. CLAWS (the Constituent Likelihood Automatic Word-tagging System) tagger was constructed by the UCREL of Lancaster University in the early 1980s and it is also available online. The system has an error rate of 1.5% and therefore has consistently achieved the accuracy of 96-97%. As for the English Vocabulary Profile website, the thorough data collection of learners all over the world and constant use of sources related to general English, such as examination vocabulary lists, classroom materials and a wide range of course books, EVP's reliability and validity is conclusive.

Summary

In this chapter, the writer had discussed the research design, which was conducted in a mix method of both quantitative and qualitative approach. She specified the decisions regarding to the choices of using both approaches, participants, and the methodology. Additionally, two kinds of data analyses were shown. This chapter ended in the description about the procedures of the data analysis.

FINDINGS

The result collected through CLAWS Tagger C5 set analysis and and AntConc application is used to count the number of and the types of nouns applied by the students in their poetry writing which was tagged by the code *NN*. This code indicates Nouns according to CLAWS Tagger Tagset5. The frequency of Nouns in Claws Tagger is divided into four categories which are Neutral Noun, Singular Noun, Plural Noun and Proper Noun as illustrated in Figure 1. Although the lesson covers many other types of noun (as described above), the CLAWS Tagger Tagset5 only included four tags for noun which are, *neutral, singular, plural and proper noun*. While in Tagset7, although the tag has been upgraded with, for instance *singular noun for direction, following noun of title* and so on, but the tags still do not cover the types on noun taught in the lesson. Therefore, this study will only discuss the four types of noun tagged namely *neutral noun, singular noun, plural noun and proper noun.* The frequency of Noun form in the students' poetry writing is shown in Table 1.

TABLE 1. The frequency of Nouns

Types of Nouns/ Frequency	Neutral Nouns	Singular Nouns	Plural Nouns	Proper Nouns
	0	83	20	7

The frequency and the types of nouns appears in their poems are both compiled and tabulated in the illustrated figures along with the discussion.

DISCUSSION

The discussion of this study will answer the research questions mentioned in Chapter 1.
    (a) What is the frequency level of the nouns used by undergraduate students in their poetry writing?
    (b) What category of nouns is mainly used by the students?
As for the frequency level of the nouns forms in the student's writing, the number is highest compared to other tagset for instances the Personal Pronouns (PNP), Adverb (AV) and Adjectives (AJ). Among the examples of the Nouns used by the undergraduates are as illustrated in Table 2 below.

TABLE 2. Examples of Nouns used by the undergraduates

No.	Types of Nouns	Examples
1.	Singular	Teacher, candle, heaven, chef
2.	Plural	Stars, words, friends, trees
3.	Proper	Justin Bieber, Selena Gomez

Most of the nouns used are listed in Paul Nation's first 10 000 headwords which is considered as 'extremely useful for learners' (Schmitt & Schmitt, 2011). For instance, the word 'smile', 'nature' and 'friend' used by the students can be found in Nation's first 10 000 list. Furthermore, for the type of nouns with the highest frequency, the analysis shows that all 10 students mostly use Singular Nouns (Figure 2) and none of the students use Neutral Nouns for example 'parent' or 'child' in their writing. This is similar to the research done by Kaur & Sarimah (2011) that mentions, "The overall indication shows that the ratio is very high for the singular noun form compared to the other noun forms" (p. 7). The excessive use of Singular noun is understandable as poetry has always been considered as 'intimate writing' where the author gets to pour his or her heart out while writing. As the students has been exposed to literary components in their secondary school, and it is among the objectives of the literature programme for students to be able to give their personal responses (Sukatan Pelajaran Kesusasteraan Dalam Bahasa Inggeris, 1990, p. 8-9), writing the poems give them freedom to voice out feelings and emotions. And as Hyland (2000, p. 78) notes, writers need to "project an insider ethos".

Meanwhile, for all 10 writing, the second highest frequency comes from plural noun and the total difference between singular and plural noun is 63. According to Laidlaw (1987, p. 60), plural noun is a noun that names more than one person, place or thing. The big gap in frequency of singular and plural noun is possibly contributed by the rule governing plural noun. For instance, to form a plural noun from the word 'teacher', the students must apply the rule of adding -s to the word and comes up with 'teachers'. Ade Irma Suryani (2011, p. 3) states, 'they (students) have difficulties to form plural in English'.

As for Proper Noun, the count is 7 occurrences in the poems which reflects low frequency of usage. Nation and Wang (1999, p. 358) postulate that 'proper nouns could be easily understood' while years before, Hirsh and Nation (1992, p. 691) believe that proper noun does not require previous learning, 'their form (initial capital letter) and their function clearly signal they are proper nouns'. However, one should understand that to be able to use proper noun, one has to have sufficient knowledge on the subject matter, for example Justin Bieber. Brown (2010, p. 357) argues that unfamiliar proper nouns can interrupt the flow of reading,

therefore correct use of proper noun as portrayed by these students show impeccable general knowledge on the central character of the poem.

Moreover, understanding which part of speech items that is mostly frequent in student's writing is crucial as it can correct many recurring problems in the writing of English language learners. As Sedita (2005, p. 1) mentions, 'Students' word knowledge is linked strongly to academic success because students who have large vocabularies can understand new ideas and concepts more quickly than students with limited vocabularies.'. However, to teach all the vocabulary to the students is not only unrealistic, there is also barrier of culture and education level to some students. Some vocabulary requires exposure on English culture for the student to be able to grasp the meaning for instance words like 'chimney', 'cart', or 'horse stable'. However, educators could always go for relational vocabulary to explain to the students the meaning of these words such as 'cow pen' for 'horse stable'. Relational vocabulary will help students associate the unfamiliar vocabulary with their culture and thus grasp the meaning of the word. Sufficient knowledge in identifying and utilizing Noun forms will help students to go further in understanding other parts of speech i.e. Adjectives as Adjectives describes a Noun. Thus, the words should be cautiously and continuously highlighted during language learning in order to achieve the appropriate level of fluency and comprehension.

The high frequency level of Nouns indicates that other than having ample vocabulary knowledge to create their poems, the students are able to identify the rules of subject-verb agreement successfully. This is true if further and thorough examination on sentences analysis is carried out. Some examples of correct subject-verb agreement rules applied by the students in the poems are *friends are, your sacrifices have, friendship is,* and *nature has.* As Kaur & Sarimah (2011) postulate, "the word choices are reflective of the learners' vocabulary knowledge and the ability to indicate the number element of the subjects and the objects of the sentences" (p. 6).

Referring to Table 3, it is the summary of the CEFR level of English vocabulary that the students acquire. Mostly, the words that the students used in their poems fall in A1 level, which is considered as Basic User (Breakthrough). Among the words are *teacher, love, friend,* and *river.* There are three characteristics of the basic user based on the CEFR Global Scale, which are:

    • Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type.
    • Can introduce him/ herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has.
    • Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

(Cambridge University Press, 2013)

While A2 and B1 level share average number of 14 and 19 words that the students use, it actually shows that Malaysian undergraduates level of English is at par the global standard. Looking at the text of the poetry, A2 and B1 words appear in all 10 poems. However, there is no outstanding numbers of A2 and B1 vocabulary for each texts; meaning that all students have an average of 2 to 3 words of A2 and/ or B1. Although measuring one's vocabulary size is difficult as researcher has to decide the counting unit that they are focusing on i.e word family or lemma (Gyllstad, 2013, p. 19), we can assume that once a student has understood the inflectional system of a language (for example, house - houses - housing), they can straightforwardly link different form of noun together. As A2 and B1 is the level that undergraduates should be at, while putting effort to achieve B2 for their exit university level,

it is clear that with the increased of the vocabulary, their vocabulary size will likely extend. Words like *experience, friendship, chance,* and *truth* appears in their writing. Therefore, the noun used is appropriate for undergraduate student level and some of them can be seen utilising high level words such as *faith, hardship* and *sacrifice.*

TABLE 3. Total number of Nouns based on CEFR level

CEFR level	A1	A2	B1	B2	C1	C2
No. of Nouns	27	14	19	9	4	1

CONCLUSION

This paper is written in order to examine a particular linguistics element which is Noun. Data from student's writing is analysed to discover the pattern of Noun utilisation form which apparently shows four types of Nouns being used, *singular, plural, neutral* and *proper*. The frequency of Nouns used by the students is the highest compared to other part-of speech of English. Thus, it can be safely said that the answer to research questions were found although the sample used is limited and may not be generalized to a bigger sample. Apart from that, the findings should also shed some lights into understanding frequency, POS tagging and CACA methodology. Furthermore, the appropriate level of CEFR vocabulary that the students portray is somehow encouraging and it brings positive perspective towards the level of English among Malaysian undergraduates. The local graduates are doomed to perception such as 'notably lack command of the English language' (Lim & Mardziah, 2014), 'poor grasp of the English language among Malaysian university graduates…' (Chin, Ting & Yeo, 2016). Thus, by exploring student's choice of words, noun in particular, it is an eye-opening to realising Malaysian undergraduates' potential and level of English. Although some might say, the assumption is invalid due to small number of sample, this study should be a starting point to celebrate more research in this field.

REFERENCES

Ade Irma Suryani. (2011). An analysis of the students' error in learning plural forms of nouns.

Baker, M. (2003). On the loci of agreement: inversion constructions in Mapudungun. Proceedings of NELS 33.

Baron, A., Rayson, P., & Archer, D. (2009). Word frequency and keyword statistics in historical corpus linguistics. Anglistik: International Journal of English Studies, 20, 1, 41-67.

Black, N. (1994). Why we need qualitative research. Journal of Epidemiology and Community Health. Issue 48. (p. 425 & 426).

Cambridge University Press. (2013). Introductory Guide to the Common European Framework of Reference (CEFR) for English Language teachers.

Chin, V., Ting, H., & Yeo, J. (2016). Investigating English language anxiety among UiTM Sarawak undergraduates. Journal of Creative Practices in Language Learning and Teaching (CPLT), 4, 50-62.

Gries, S. (2009). What is Corpus Linguistics? Journal of Language Linguistics Compass, 3, 5, 1225-1241.

Goldin-Meadow, Susan, Cynthia Butcher, Carolyn Mylander, & Mark Dodge (1994). Nouns and verbs in a self-styled gesture system: What's in a name? Cognitive Psychology, 27, 259–319.

Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions (pp. 11-28). Journal of L2 Vocabulary acquisition, knowledgement, and use: New perspectives on on assessment and corpus analysis.

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? Reading in a Foreign Language, 8, 689–696.

Kaur, M. & Sarimah Shamsudin. (2011). Extracting noun forms: A lesson learnt. International Journal of Language Studies 5(4): 19–32.

Laidlaw. (1987). Laidlaw English. Illinois: Laidlaw Brothers Publisher.

Lim, H.L., & Mardziah Budin. (2014). Investigating the relationship between English language anxiety and the achievement of school based oral English test among Malaysian form four students. International Journal of Learning, Teaching and Educational Research, 2(1), 67-79.

Nation, I.S.P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Nation, I. S. P. (2007). The four strands. Innovation in language learning and teaching, 1, 1–12.

Nation, I. S. P., & Wang, M. (1999). Graded readers and vocabulary. Reading in a Foreign Language, 12, 1999.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. Journal of Mixed Methods Research, 3, 208-222.

Sapir, E. (1921). Language: An introduction to the study of speech. New York: Harcourt Brace Jovanovich.

Schmitt, D., & Schmitt, N. (2011). Focus on vocabulary 2: Mastering the academic word list. White Plains, NY: Pearson Longman.

Sedita, J,. (2005). Effective vocabulary instruction. Journal of Insights on Learning Disabilities, 2, 1, 33-45.

Tognini-Bonelli, E. (2001): Corpus Linguistics at Work. (Studies in Corpus Linguistics 6). Amsterdam/ Philadelphia.

Wehmeier, S. (2000). Oxford Advanced Learner's Dictionary of Current English. (6th Ed.). University Press: Oxford.

AUTHORS

Nur Baiti Ismail Shauki is a PhD student in Applied Linguistics in Universiti Utara Malaysia. She is currently teaching in Universiti Malaysia Kelantan. Her research interests are computer-assisted corpus (CACA) and language learning and teaching.

Manvender Kaur A/p Sarjit Singh (PhD) is a lecturer in School of Education and Modern Languages in Universiti Utara Malaysia. Her research niches are corpus linguistics and discourse.