Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Using Technologised Computational Corpus-Driven Linguistics Study on the Vocabulary Uses Among Advanced Malaysian Upper Primary School English as a Second Language Learners (ESL) in Northern Region

Wong Wei Lun[1], Mazura Mastura Muhammad[1,*], Warid Mihat[2], Mairas Abdul Rahman[3], Mohamad Syafiq Ya Shak[4], Lee Mei Chew[5]

[1] Department of English Language and Literature, Faculty of Languages and Communications, Sultan Idris Education University, 35900 Tanjong Malim, Malaysia
[2] Academy of Language Studies, University Teknologi MARA Kelantan Branch, 18500 Machang, Malaysia
[3] Faculty of Languages and Communication, University Sultan Zainal Abidin, 21300 Kuala Nerus, Malaysia
[4] Academy of Language Studies, University Teknologi MARA Perak Branch, 32610 Seri Iskandar, Malaysia
[5] Department of Science and Mathematics, Faculty of Science and Mathematics, Sultan Idris Education University, 35900 Tanjong Malim, Malaysia

ARTICLE INFO

Article history:
Received 24 January 2023
Received in revised form 1 June 2023
Accepted 7 June 2023
Available online 23 June 2023

Keywords:
Corpus-driven; vocabulary; learner corpus; ESL learners

ABSTRACT

The vocabulary used by advanced Malaysian upper primary school learners is unknown after the Covid-19 pandemic, according to a corpus-driven study. This study primarily uses data from a learner corpus. This main purpose of this paper is to fill in this knowledge gap by examining the vocabulary usage in extended writing by advanced Malaysian upper primary school learners in the Northern Region. The study uses a corpus-driven research strategy and a quantitative methodology. Some 160 advanced upper primary school learners from the Northern Region, Malaysia were included in the study. They attended eight national primary schools with strong English programmes in the Northern Region. Purposive sampling technique was used to select the samples. This study's main corpus-driven toolkit (LancsBox) was used to analyse the data. A total of 46,185 tokens from 160 extended writing tasks were analysed. The word frequencies were counted both raw and normalised. Results showed that 160 advanced Malaysian upper primary school learners in West Malaysia used the frequently and preferably in extended writing followed by the, and vocabulary be is the twentieth salient vocabulary over other vocabulary such as to, and, a, and I. Depending on their perspectives, one may argue that they described their experiences in the first person. These findings imply that advanced Malaysian upper primary school learners use to, and, a, of, in, is, it, that, for, as and within their extended writing. The findings provide important knowledge and information for ESL teachers to improve the writing grammatical proficiency, especially for extended writing of ESL learners.

* *Corresponding author.*
E-mail address: mazura@fbk.upsi.edu.my

https://doi.org/10.37934/araset.31.1.298314

1. Introduction

1.1 Research Background

Globally, corpus linguistics has grown into a major area of research [1]. Despite this, it has not been frequently subjected to rigorous exploration in Malaysia. Additionally, corpus-driven research focusing on upper primary school learners is still scarce. In other words, there is essentially a huge research gap for more local research on corpus linguistics [2]. In Malaysia, English is taught as a second language [3,4]and the English subject is compulsory for all students in national primary schools, regardless of their ethnicity or culture [5]. Furthermore, the Common European Framework of Reference (CEFR) serves as the key framework for assessment, teaching and learning English as a second language [6] and some of the success statements in CEFR are students can learn essay writing, vocabulary and phrases. The learning criteria are as follows: (a) 4.2.4 Use appropriate sentences to describe people, places, and objects; and (b) 4.3.2 Spell various high- frequency words accurately in autonomous writing. To meet these learning objectives, vocabulary and phrases are critical components of essay writing for language formation. When compared to other abilities such as listening, speaking, and reading, writing is considered the most challenging since it demands ESL students to be able to acquire both vocabulary and grammar concepts and use them to compose sentences correctly and effectively [7].Teaching writing skills to ESL students is thus undoubtedly difficult [8-10].

As afore mentioned, English is one of the compulsory subjects in Malaysian primary schools with the objectives of assisting the learners to master the language and the skills. The introduction of the CEFR emphasised the importance of students' ability to write descriptive essays and to enable the students to compose clear and cohesive descriptive essays requires sufficient vocabulary and phrases. However, past research shows that many primary school learners struggled to employ suitable vocabulary and phrases to convey clear information in essays [11]. Moreover, research has shown that English teachers have encountered problems and are unsure of the appropriate phrases to teach to make a good essay [12,13] reiterates the lack of corpus research on the use of phrases for school learners and English teachers. [14] investigated the most prevalent errors made by Malaysian Chinese primary year 6 ESL students when writing in English as well as the degree to which the first language, Mandarin, affected ESL writing. English is written using alphabetic scripts, while Mandarin utilises a logographic writing system. Chinese ESL students find it difficult to master the English language because of the differences between the two language systems. A qualitative research methodology was employed which involved writing activities, with 15 ESL learners from primary year six selected as participants. The research found that the areas with the highest and most noticeable grammar mistakes were tenses, object pronouns, plurals, auxiliary verbs, prepositions, and articles. It was discovered that these inaccuracies were brought on by the influence of the mother tongue through direct translation from the Mandarin language. In conclusion, mistakes will inevitably occur, but they are essential for ESL students to advance in their English writing ability. Grammar is the key linguistic function of the English language that will be impacted by language transfer. For ESL learners, grammar is thought to be the most beneficial part of the language. Before the Malay language was replaced in public secondary schools in 1981, the language was widely utilised as a medium of teaching, hence it is essential for students in Malaysia to have a strong grasp of English grammar [7].

In another study carried out by [15] on Malaysian primary students' problems with narrative writing, findings showed that the learners' L1 had a significant impact on their L2. Other elements including the students' earlier experiences and helpful and constructive criticism on their writing could have contributed to their struggles with narrative writing. [16] agreed that writing is complex,

demanding, and requires organisation skills, writing mechanics, correctness in word choice, and appropriate knowledge of syntax and grammar; it is problematic for ESL learners. [17] concur that ESL learners struggle to create basic sentences, and the majority of them have difficulty creating compound and complicated sentences that contain dependent and independent clauses. Along with spelling, L1 interference, and language use (grammar), these are the main difficulties faced by ESL students [18]. However, previous studies have mostly focused on comprehending the difficulties that secondary and tertiary students have in learning. Very few studies explored vocabulary use among advanced learners in selected schools in Malaysia. Therefore, this study aims at collecting and analysing vocabulary use among advanced Malaysian upper primary school learners of English. The emphasis will be on writing ability, especially vocabulary components. The present study's findings aimed at contributing to the current Malaysian expertise in CL research and the use of vocabulary in extended writing through examining advanced Malaysian upper primary school learners' extended writing. This study will answer the following research question: What are the differences and/or similarities in the use of English vocabulary in the extended writing produced by advanced learners in upper primary schools in the northern region of Malaysia?

1.2 Literature Review and Previous Studies on Learner Corpora

Vocabulary is a critical component of language acquisition. School learners must improve their vocabulary to communicate their thoughts. According to [19], vocabulary is defined as school learners' comprehension of oral and written words, including conceptual knowledge of the terms beyond their straightforward dictionary definitions. They emphasised that vocabulary acquisition is a continual process in which school learners make links to other words, study instances of related words, and eventually use the vocabulary correctly and appropriately within the context of the sentence. Furthermore, [20] defined vocabulary as a language's words, including single words, sentences, or chunks that convey meaning. In this study, the term vocabulary refers to advanced Malaysian upper primary school learners' words and phrases utilised in extended writing.

Based on the social constructivism perspective, humans view the world through eyes shaped and formed by experiences and interactions. Reality is unique to each individual and contextualised by the social context in which activities occur. If this is true, then learning is not universal, and learners will react differently to classroom surroundings, classes, and professors based on their prior experiences and previously constructed knowledge. Learners in a classroom bring a unique reality, or vision of reality, to their study, and together they will develop new understanding. With numerous alternate realities colliding in a common space, schools can be either rich or poor, depending on the atmosphere created. Teachers play a critical role in the design of a class and their responsiveness to their learners from a social constructivist perspective on learning. The school cannot serve as a preparation for social life unless it replicates the typical conditions of social life within itself (Dewey, 1909). This also means that if education's goal is to prepare learners for life in the world, authentic learning is critical, ensuring that what we do in the classroom reflects real life.

Generally, a corpus-driven study employs empirical corpus data from which language features arise naturally through data analysis. The availability of prominent and representative corpora and the incorporation of computational software enables the investigation of linguistic variation from various angles. The prospective corpus is used to generate linguistic categories that have not been recognised by corpus linguistics or scholars, as the findings are intended to be exhaustive in terms of corpus evidence [21]. The linguistics categories are formed systematically from the recurrent patterns and frequency distributions when language is used in context [21]. According to Love [22], certain corpus-driven studies emphasised the importance of frequency evidence, particularly when

studying lexical bundles, while not prioritising frequency in analysing grammar patterns. Despite this contrast, it is assumed that the primary goal of a corpus-driven study is to discover novel language features inside a corpus inductively.

Before CL, linguists studied grammar based on the intuition of native speakers. The corpus has made a considerable contribution to linguists, researchers, teachers or students to use natural language data to show and understand how language works. They can analyse many aspects of language such as vocabulary, including frequency, collocation and grammar, grammatical rules in grammar books, describing language, and displaying qualitative and quantitative analysis. In addition, they can compile a corpus of a specific type of language and find commonly used words and expressions. They can even compare students' written and spoken language to determine the types of mistakes they make [23].

Undeniably, the advantage of CL is providing the research to give quantitative manner to linguistic features through systematic statistical measures such as computer software, which is laborious if performed through manual steps. [24] elaborates on CL based on computer technology by stating that it is an advanced emerging methodology to analyse language and a new philosophical approach for language because it is a robust technological tool that has made CL possible. [25] added that to use appropriate software to measure language, tools for indirect observation such as collocates, query languages, aligners, concordances and parsers are required.

In recent years, CL has been employed in research on vocabulary. It includes using corpora in teaching and learning a language. For instance, [26] has shown corpora to teach a language. In 2020, the vocabulary programme he designed allowed language teachers to arrange their corpus by having 32 various texts with different lengths to identify words prevalent in them. [27] used learners corpora to investigate the use of boys' and girls' vocabulary in native and non-native contexts. Both of the researchers applied WordSmith Tools 5.0 to compare critical verbs that are key in domains related. In 2020, Love compared the vocabulary used by speakers of different genders, ages and social groups to study a spoken conventional component of BNC which consisted of 4.5 million words. From the gender aspect, 25 most essential words characteristic of both genders were found. In the findings, boys tend to use more cursing words while females prefer to use more feminine pronouns and first-person pronouns.

1.3 Corpus Linguistics (CL) in Malaysia

Regarding past studies conducted [28-32], Malaysians are becoming more interested in CL, especially in learner corpora. Other research areas of language are included, such as Malay linguistics by [33], and English for specific purposes by [34]. These studies employed a different set of methodological approaches to investigate various linguistic features, namely grammar and semantics. Interestingly, English was not the only language explored as the Malay Language was targeted for research. On the other hand, these studies have not examined vocabulary and phrases used by young learners in extended writing.

[35] conducted research similar to the present study, which investigated the types of discourse connectives used by 32 secondary students in narrative essays. In addition, interviews were carried out with ten selected students to explore how they perceived the use and meaning of some discourse connectives in their texts. The researcher examined 96 narrative essays and listed out the discourse connectives identified in them. Based on the preliminary findings, she found that the students tend to use different discourse connectives in their essays. However, they were alert that there were many other types of discourse connectives available in English. She added that students were innovative in applying discourse connectives in essays instead of strictly following the related frameworks.

Linguists were made aware that students were able to become accomplished English users despite imitating native speakers in using the discourse connectives for essay writing. Although the mentioned research examined student use of discourse connectives in narrative essays, it is meaningful to examine from the perspectives of advanced Malaysian upper primary school learners their use of vocabulary in extended writing. In this present study, learner corpora were the corpora chosen and collected since the data collected were authentic extended writing from advanced Malaysian upper primary school learners who learn ESL. It fulfils second language acquisition because they learn ESL in Malaysia [5]. The number of extended writings collected was around 160 essays. Hence, learner corpora were suitable as the extensive data were stored electronically and analysed through suitable software and applications.

This study was conducted using a corpus-driven analysis. The corpus obtained was learner corpora, meaning the results were based on actual data gathered during the investigation rather than on a reference corpus as in corpus-based analysis. The linguistic component examined was vocabulary used by 160 advanced Malaysian primary school learners in extended writing. Hence, the extended writing was used automatically in this study to gather learner corpora. To begin, LancsBox was utilised to extract salient vocabulary (functional & content words) from the extended writing. The research focused on functional and content words since functional words may not give meaningful results. Moreover, to illustrate, the log-likelihood calculator was utilised for the log-likelihood values while comparing the salient vocabulary (functional & content) to the L-O-B reference corpus.

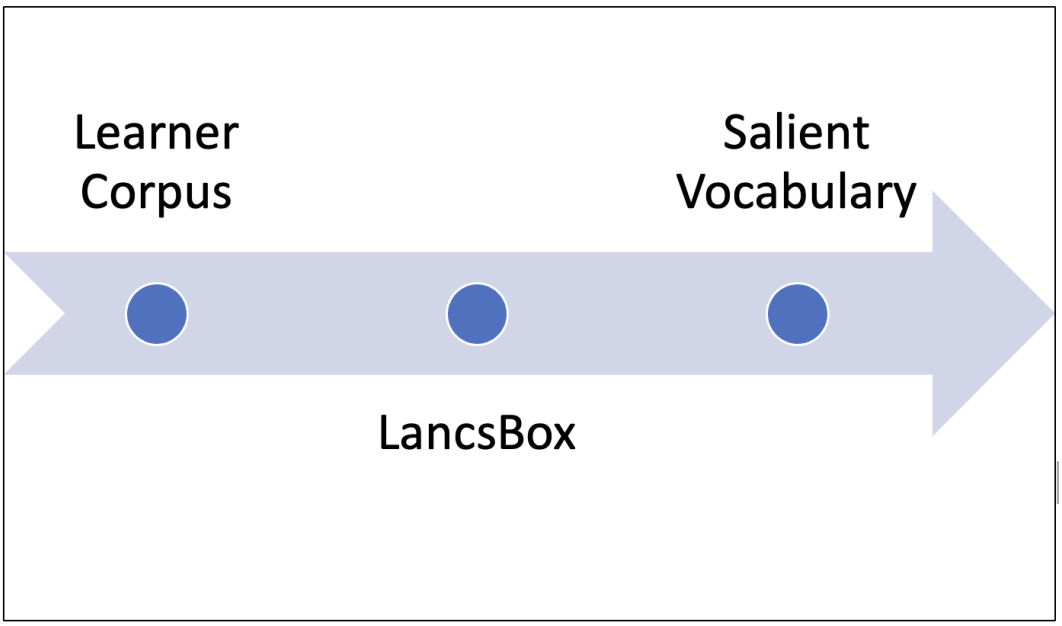

Fig. 1. Research conceptual framework

2. Methodology

The study uses a corpus-driven research strategy and a quantitative methodology. The current study aimed at analysing, comparing and contrasting the top 20 salient vocabulary (functional & content) used among advanced Malaysian upper primary school learners of English. The study focused on vocabulary used by 160 advanced Malaysian primary school learners in their extended writings. To establish a comparable scenario for collecting written texts in the four national primary schools and contexts, it was determined that collection methods needed to be as similar as feasible.

The research site chosen was the northern states from Malaysia involving 4 national primary schools with high English performance located in Malaysia. The national primary schools chosen were Sekolah Kebangsaan (henceforth, SK) and Sekolah Kebangsaan Jenis Kebangsaan Cina (henceforth, SJKC) based on the data provided by officers from the District Education Office and State Education Department.

2.1 Participants of The Study

Based on Table 1, Peninsular Malaysia specifically the northern region was chosen as the research site. The study included only the capitals of each state because the target participants were advanced Malaysian upper primary school learners enrolled in national primary schools with high English proficiency, as determined by district education offices or state education departments. Only forty advanced primary school learners from each state were chosen to complete the extended writing assignment. As a result, the number of advanced primary school learners and extended writing were generated concurrently. Due to the nature of the extended writing, all advanced primary school learners were given the flexibility and opportunity to write on any subject they desired.

Table 1
Demographic data of AMUPSLC

East/West Malaysia	Region	Capital	No. of Learners/Extended Writing	No. of Tokens	No of Tokens per Region
		Perlis	40	9,855	
Peninsular	Northern	Alor Setar	40	8,260	46,185
Malaysia	Region	Georgetown	40	10,392	
		Ipoh	40	17,678	
Total	1	4	160	46,185	46,185

2.2 Data Collection

The following sections outline the main design concerns and criteria for the writing task

i. each advanced upper primary school learner writes one extended writing of their chosen topic (without any prompts given by the teacher)
ii. face-to-face or online classroom environment
iii. no word limits
iv. 30 minutes length of time (flexible)
v. advanced upper primary school learners are not permitted to use dictionaries or other reference source.

When the writing task was assigned, the English teachers gave the following directions to the pupils: Today, you will write on whatever subject you want. Research participants were given a minimum of 30 minutes to complete a writing task in class or at home through online learning. It should be noted that time and word limitations were not strictly enforced since upper primary school learners were given flexibility if they needed more time or wrote over the word limit.

The handwritten writings of AMUPSLC were first typed and stored as doc/docx files. All files were separated by states into thirteen different folders. Subsequently, they were further separated by regions into five different folders. Spelling mistakes were rectified during transcription. Because the purpose of this current study was not to determine how much advanced Malaysian primary school

learners' extended writing varies from Standard English, the researcher did not analyse spelling, grammar, or punctuation mistakes. Rather, the researcher was interested in studying their choices of vocabulary. Therefore, some degree of standardisation needed to be included in the transcriptions to aid in identifying trends and patterns in phrase usage. For instance, the researcher wanted to be able to detect all instances of the term wake up, considering alternative spellings (wak up, wok up) as synonymous for the sake of this study. While converting the handwritten texts to computer format, the researcher encountered numerous typical difficulties when analysing the extended writing. Nonetheless, there is an inherent problem in transcribing their writing, based on unreadable texts and concerns about spelling and grammaticality mistakes. However, even if school learners sometimes made spelling or grammatical mistakes, it was usually simple to decipher what they meant to write.

To answer research question one which focused on emphasising the salient vocabulary (functional & content) the researcher chose to use a variety of automated techniques to establish significant variances and similarities. LancsBox is a suite of corpus software developed by Lancaster University. It includes tools for doing several types of linguistic analysis, including WordList, Keywords, concordances and log-likelihood calculator. The next step was to separate the frequency lists' function and content terms. This was accomplished manually by scanning each list for the first twenty salient vocabulary (functional & content). The top twenty salient vocabulary (functional & content) were quantified in this study using frequency lists. What makes the frequency lists usage so important is that they disclose information about the corpus. With reference to Morato *et al.,* (2021), frequency lists constitute a fundamental kind of frequency information, but they also offer critical information to supplement the findings of thorough concordance studies. Frequency lists may be used to characterise the corpora from which they were derived. The researcher chose the top twenty salient vocabulary (functional & content). Typically, the top-ranking vocabulary (functional & content) on the list are functional words.

The keyword analysis employs two main corpuses namely AMUPSLC and L-O-B corpus, with the later serving as the reference corpus, which is compared using Lancsbox. The frequency lists for each state were compared. The comparison findings include terms that are 'key' in one file but not in the other. After uploading the texts in the corpus to calculate the frequencies, these values were normalised to an incidence rate per 1,000 words of text. A review of concordance lines may aid in understanding the situations in which advanced Malaysian upper primary school learners employ this vocabulary (functional & content).

2.3 Data Analysis

To answer research question one which focused on emphasising the salient vocabulary (functional & content) the researcher chose to use a variety of automated techniques to establish significant variances and similarities. LancsBox is a suite of corpus software developed by Lancaster University. The process was outlined in the previous section. The top twenty salient vocabulary (functional & content) were quantified in this study using frequency lists. What makes the usage of frequency lists so important is that they disclose information about the corpus. With reference to [36], frequency lists constitute a fundamental kind of frequency information, but they also offer critical information to supplement the findings of thorough concordance studies. Frequency lists may be used to characterise the corpora from which they were derived. The researcher chose the top twenty salient vocabulary (functional & content). Typically, the top-ranking vocabulary (functional & content) on the list are functional words.

3. Results

3.1 Functional Words

The following part summarises the top twenty salient functional words in Peninsular Malaysia (Northern Region). Next, their log-likelihood values are analysed and the functional words are examined and analysed together with the concordances.

3.2 Salient Functional Words: Peninsular Malaysia

In the Peninsular Malaysia sub corpus, there are 480 extended writing with 132,614 tokens. First and foremost, a total of 160 extended writing was imported into LancsBox for data analysis from the sub corpus of the Northern Region. To determine the salient functional words, a total of 46,185 tokens were analysed. Through the acquisition of frequency lists, twenty salient functional words were found. Table 2 presents the twenty salient functional words in terms of raw frequency (henceforth, RF), normalised frequency (henceforth, NF) and percentage of occurrence (%) in each of the extended writing. The column N refers to the number of essays where these words appear, for example, *the* is found in 160 written texts in the frequency list.

Based on Table 2, 160 advanced Malaysian upper primary school learners in Peninsular Malaysia used *the* frequently and preferably in extended writing. The RF is 2417.00, which looks to be the highest of the remaining 19 functional words. Additionally, it is the sole vocabulary found in all 160 extended writing analysed. In contrast to *the*, *be* is the twentieth salient vocabulary, with a RF of 277.00. In extended writing, only 92 advanced Malaysian upper primary school learners utilised the vocabulary *be*. It might be deduced that *be* was not given priority over other vocabulary such as *to*, *and*, *a*, and *I*. Surprisingly, vocabulary *I* is the fifth salient functional word, appearing in just 95 of 160 extended writing. The percentage of N is 59.38. It reveals that 95 advanced Malaysian upper primary school learners frequently utilised the vocabulary *I* in their extended writing. Depending on their perspectives, one may argue that they described their experiences in the first person.

Table 2
Salient Functional Words of West Malaysia: Northern Region

Rank	Word	RF	NF	%	N	N%
1	the	2,417.00	523	5.23	160	100
2	to	1,658.00	359	3.59	159	99.38
3	and	1,400.00	303	3.03	157	98.13
4	a	1,050.00	227	2.27	155	96.88
5	I	888.00	192	1.92	95	59.38
6	of	752.00	163	1.63	138	86.25
7	we	699.00	151	1.51	90	56.25
8	in	681.00	147	1.47	142	88.75
9	is	607.00	131	1.31	116	72.50
10	was	603.00	131	1.31	93	58.13
11	my	590.00	128	1.28	93	58.13
12	it	541.00	117	1.17	124	77.50
13	that	476.00	103	1.03	123	76.88
14	for	367.00	79	0.79	123	76.88
15	as	296.00	64	0.64	103	64.38
16	you	290.00	63	0.63	46	28.75
17	are	290.00	63	0.63	98	61.25
18	they	286.00	62	0.62	82	51.25
19	with	278.00	60	0.60	102	63.75
20	be	277.00	60	0.60	92	57.50

Note: The raw frequencies have been normalized to as per ten thousand words for uniform representation. The formula applied is (N_{RF} x 10^4) / T = N_{NF} whereby, N_R = value of the raw frequency of the salient vocabulary, T = total tokens of the sub corpus (46,185) and N_{NF} = value of the normalised frequency. The figures for percentage distribution are rounded off to the nearest two decimal digits. The formula applied is (N_{NF} / 10^4) x 100% = N_{PD}%. whereby, N_{NF} = value of the normalised frequency, N_{PD} = value of the percentage of occurrence (%).

Similarly, the vocabulary *my* and *was* were used in conjunction with *I*. *My* is the eleventh salient functional word, while *was* is the tenth salient vocabulary. Their RFs are 590.00 (my) and 603.00 (was) respectively. However, *my* and *was* are discovered in 93 extended writing. It is estimated that 93 advanced Malaysian upper primary school learners regularly used *my* and *was* in their essays. Furthermore, it is suggested that 93 extended writing were produced in the past tense in order to describe prior experiences. Logically, it is reasonable to suppose that advanced Malaysian upper primary school learners in the Northern Region prefer to write extended writing about previous experiences in the first person perspective using *I*, *my* and *was*.

Other vocabulary is discovered in numerous extended writings (≥ 100) due to the findings. They are the vocabulary from the second, third, fourth, sixth, eighth, ninth, twelve, thirteenth, fourteenth, fifteenth, and nineteenth ranks. *In* has a RF of 681.00, but it is found in 142 extended writing, which is more than the sixth functional word *of*, with just 138 extended writing examples. Following that, *is* has a RF of 607.00; advanced Malaysian upper primary school learners used it in total of 116 extended writing. Other functional words such as *it* (541.00), *that* (476.00), *for* (367.00) and *as* (296.00) with lesser RF as compared to *is* (607.00) seem to appear more in extended writing by advanced Malaysian upper primary school learners which are 124, 123, 123 and 103 extended writing respectively. These findings imply that advanced Malaysian upper primary school learners use *to*, *and*, *a*, *of*, *in*, *is*, *it*, *that*, *for*, *as* and *with*in their extended writing to improve the organisation and explanation of essays through the use of pronouns, prepositions, linkers, and conjunctions.

On the other hand, the functional words which are not identified in most of the extended writing (≤ 100) are *I*, *we*, *was*, *my*, *you*, *are*, they and *be*. Their RFs are 888.00, 699.00, 603.00, 590.00, 290.00, 290.00, 286.00 and 277.00 respectively. Although they might have a relatively high frequency compared to other functional words, the number of extended writings using them is less. Only 46 to 98 extended writing have used these functional words. Hence, it is derived that most of the advanced Malaysian upper primary school learners in the Northern Region did not prefer to use these functional words and preferred to describe their essays in the third person instead of the first person.

With reference to Table 3, 14 salient functional words demonstrate significance due to their high G^2 value of $\geq$15.13 at $p < 0.0001$. The salient functional words are *the* (233.00), *to* (47.91), *a* (25.59), *I* (681.32), *of* (782.36), *we* (373.32), *in* (139.31), *was* (198.52), *my* (813.85), *it* (99.70), *for* (67.91), *you* (196.44), *they* (41.52), and *be* (42.50). The NF has little effect on the G^2, as certain functional words, such as *to*, have a G^2 of only 47.91. Simultaneously, *and* has an NF of 303 but only a G^2 of 2.41. Each of these 14 salient functional words is important compared to the L-O-B reference corpus.

Table 3
Log-Likelihood Values of Northern Region vs. L-O-B

Subcorpus (Northern Region)			vs. L-O-B	
Rank	Word	NF	G^2	Sig. Level
1	the	523	233.00	< 0.0001
2	to	359	47.91	< 0.0001
3	and	303	2.41	< 0.1
4	a	227	25.59	< 0.0001
5	I	192	681.32	< 0.0001
6	of	163	782.36	< 0.0001
7	we	151	373.32	< 0.0001
8	in	147	139.21	< 0.0001
9	is	131	11.35	< 0.001
10	was	131	198.52	<0.0001
11	my	128	813.85	< 0.0001
12	it	117	99.70	<0.0001
13	that	103	0.36	< 0.1
14	for	79	67.91	< 0.0001
15	as	64	2.97	< 0.1
16	you	63	196.44	< 0.0001
17	are	63	0.14	< 0.1
18	they	62	41.52	< 0.0001
19	with	60	9.26	< 0.01
20	be	60	42.50	< 0.0001

Note: G^2>6.63 at $p < 0.01$ or 1% level. Values of significance as given by McEnery *et al.,* (2006, p. 55): Null hypothesis, H_0 = There exists no significant association between salient vocabulary in the sub corpus of AMUPSLC with those in the reference corpora. Alternative hypothesis, H_a = There exists a significant association between the occurrences of salient vocabulary in the sub corpus of AMUPSLC with those in the reference corpora ($G^2 \geq 10.83$ at $p < 0.001$, $G^2 \geq 15.13$ at $p < 0.0001$). Similar values of significance are applied for G^2 test in AMUPSLC.

Following that, just one salient functional word bears the $G^2 \geq 10.83$ at $p < 0.001$. *Is* has a G^2 value of 11.35. Compared to the reference corpus, however, L-O-B, it is still regarded as a relatively crucial functional word. The remaining five functional words are less or non-significant compared to the reference corpus, L-O-B, because their G^2 is less than 10.83, which complies with the G^2>6.63 at $p <$ 0.01 or 1% level requirements. *With* (9.26) is the functional word with $p < 0.01$; the other four functional words with $p < 0.1$ are *and* (2.41), *that* (0.36), *as* (2.97), and *are* (0.14).

Overall, 15 salient functional words from the Northern Region's sub corpus are significant compared to the reference corpus, whereas the remaining five salient functional words are either less significant or insignificant. Additionally, one may argue that *my* is the most important functional word, with a G^2 of 813.85, while *are* is the least significant, with a G^2 of 0.14. Nonetheless, the five functional terms that are insignificant because their G^2 is less than 10.83 are moderately significant in the AMUPSLC learner corpus.

3.3 Concordances of Similar Functional Words: Peninsular Malaysia

The concordances of each similar salient functional word analysed above are presented below. As illustrated in Figure 2, the is employed as a determiner. It is used at the start of noun groups to refer to someone or something that has been stated or identified previously [37]. The use of the identifies a total of 6,768 lines of concordances (Northern: 2,417). It might be deduced that in all 2,417 lines of concordances the is used in the manner specified. Likewise, it is considered that

advanced Malaysian upper primary school learners in the Northern Region prefer to use the in their extended writing the most.

the day. During the early part of	the	day, hundreds of vehicles and pedestrians use
trucks run over me, and I feel	the	pain rushing through my body and veins.
rushing through my body and veins. During	the	night, my usage is less due to
and I am only cleaned up by	the	road sweepers the next morning and, if
only cleaned up by the road sweepers	the	next morning and, if I am lucky,
throw anything like this, throw it in	the	dustbin, not on me. Over the years,
in the dustbin, not on me. Over	the	years, due to excessive usage, I have
I just got some good news that	the	government is going to tar the road
that the government is going to tar	the	road and repair me. I will be
come and to be treated better in	the	future.
love them or thank them. One of	the	ways to appreciate others is by giving
some chocolates, balloons and many more. As	the	saying goes, action speaks louder than words.
good luck if they are sitting for	the	examination, interview and competing in any competition.
parents have decided to bring us to	the	beach. We were extremely excited because we
ages. We went there by car. During	the	ride, my sisters and I told our
a shop, it was tasty. We ate	the	snacks while watching comedy videos and movies
while watching comedy videos and movies in	the	car. Two hours later, we arrived at
car. Two hours later, we arrived at	the	beach. Without wasting any time, my sisters
and I quickly helped our parents take	the	things down from the car. After that,
our parents take the things down from	the	car. After that, my sisters, my father
changed into our swimsuits and jumped into	the	cool and refreshing ocean. We saw some
sandwiches and others. We gobbled up all	the	food in no time and chatted for
We also took some photos for memories;	the	scenery was breathtaking. We also helped clean
scenery was breathtaking. We also helped clean	the	beach because it was a bit dirty.
types of fish. We also swam in	the	sea and played beach volleyball. After that,
beach volleyball. After that, we collected seashells.	The	next day, we went to Star Shopping
Star Shopping Mall. We had lunch at	the	food court. The food was delicious. After
We had lunch at the food court.	The	food was delicious. After that, we went
Later we watch a talent competition at	the	shopping mall. After that, we went home
During	the	school holiday, my parents and I flew
I flew to Bangkok. We were over	the	moon because it was our first time.
time. Within an hour, we arrived at	the	destination. A van picked us up, and

Fig. 2. Concordances 'the': Northern Region (34 of 2,417 lines)

With regard to Figure 3, the word *to* can be employed in two ways. According to [37], the preposition *to* is used and it is also noticed before the basic form of a verb. Analysis showed that *to* is commonly employed before the base form of a verb despite of in place of a preposition. There are 1,658 concordance lines to count (Northern: 1,658). It demonstrates that advanced upper primary school learners in Peninsular Malaysia regularly used *to* preceding the basic form of a verb in their extended writing, instead of using it as a preposition.

global pandemic, it is important for us	to	adapt to the situations and find alternatives
it is important for us to adapt	to	the situations and find alternatives to make
adapt to the situations and find alternatives	to	make things work. One of the new
I believe online learning has helped me	to	improve my understanding in classes. This is
using past classes videos. It helps me	to	catch things up if anything ever comes
disadvantages—one of them was being not able	to	organise my time properly. The main reasons
reunite with our loved ones and return	to	our normal daily routine. The Covid-19 pandemic
stressful life than what it is supposed	to	be. However, life goes on. It is
life goes on. It is our duty	to	observe and adapt to the situation so
is our duty to observe and adapt	to	the situation so that we can hover
it smoothly. There are pros and cons	to	everything. One of the advantages of online
online learning is our ease of access	to	information online. We can easily search for
management and workspace. We can choose when	to	study because we are on our own.
In addition, learning at home exposes us	to	distractions such as getting interfered with by
our family members or having the urge	to	play video games. If I were to
to play video games. If I were	to	choose, classroom learning would forever be my
The environment is what makes us want	to	study. Being at home just feels like
a holiday which interferes with our mindset	to	learn. To conclude, classroom learning will always
which interferes with our mindset to learn.	To	conclude, classroom learning will always be my
especially the students. Since everyone is prohibited	to	go outside to prevent getting infected from
Since everyone is prohibited to go outside	to	prevent getting infected from Covid-19, students must
learning is a foreign way for students	to	learn. As a student, I get a
benefits me more as it is easy	to	look up useful information online. Website such
of time. Moreover, online learning allows me	to	use my laptop. Essays and reports which
Word provide a quick and easy way	to	detect errors. Therefore, I can manage my
my phone, and this resulted in me	to	divert my attention from my class. Rainy
internet connection and causes laggings. I happen	to	miss some of my classes because of
This is because I have been adapted	to	it since I was young. Classroom learning
Four students. They will give their everything	to	make sure the events will run smoothly.
of The Year that will be given	to	the student that is excellent in both
principal. The school can encourage the students	to	win the awards by organising this event
perform their best in academics and co-curriculum	to	win the awards. Other than that, the
Other than that, the school also need	to	prepare the best prize for the winner
so that the students will get excited	to	win the awards. I hope everyone will

Fig. 3. Concordances 'to': Northern Region (34 of 1,658 lines)

The concordances of *and* in four sub corpora are depicted in Figure 4. It is obvious that *and* is utilised as a conjunction by every advanced Malaysian upper primary school learner in Peninsular Malaysia. It is used to establish a connection between two or more words, groupings, or phrases [37]. A total of 1,400 concordance lines are available (Northern: 1,400). The findings are comparable to those of *the* and *to*, as advanced Malaysian upper primary school learners in the Northern Region emphasised *and* in their extended writing. One could argue that they are more adept at structuring complicated sentences.

and my group. I have already discussed	and	divided the work among all the group
we prefer to interview someone in English	and	write an article about it since one
research about how to conduct an interview	and	practice it with my family members. I
easy to get bond with the environment	and	easy to survive. My two-week holiday already
going to a picnic, visiting popular places	and	going to the waterfall. It is true
waterfall. It is true that my classmate	and	I enjoy hiking because we all have
love doing adventure activities. We have discussed	and	decided to go to Bukit Kledang for
already gone for hiking at Bukit Kledang,	and	she said it was worth it. For
we will take care of each other	and	do not let anyone go on their
in danger, you can call your parents	and	the police immediately. Next, you need to
lives on Earth. People are losing lives,	and	businesses are affected badly. There are many
do activities such as bungee jumping, parasailing	and	many more. Besides that, I also learnt
lots of people had lost their jobs	and	were facing bankruptcy problems. I appreciate that
have a cosy home to live in	and	delicious food on the table. Last but
conclude, the Covid-19 has affected the economy	and	many lives worldwide.
we've wanted to do, like bungee jumping	and	parasailing. Besides, I must be thankful for
now. Lots of people lost their jobs,	and	they couldn't provide for their families. I
appreciate that I still have a home	and	food on the table. On the other
way of life is having online classes	and	wearing masks in public places. To conclude,
has affected badly to our world's economy	and	millions of lives.
in the first place. "Eula, Amber, Jean	and	I will be moving to Fontaine. We
their adventures. "Goodbye, Diluc, Eula, Amber, Jean	and	Kaeya! Remember, like Katheryne always said, the
the traveller. Kaeya picks up his backpack	and	begins walking the opposite way as everyone
took was a few of Klee's bombs,	and	the place was gone. He arrives at
Stone Gate, the place that separates Monstadt	and	Liyue. The area was almost destroyed due
thing," Kaeya responded. "Please sign this contract,	and	you will be free to go." Kaeya
to go." Kaeya signs the given contract	and	continues making his way to the city.
it was already night. Kaeya walks around	and	finds a small cottage for rent. It
to himself. Kaeya looks around the cottage	and	sees a man about the same height
his backpack in the cottage he rented	and	begins walking to the city. He hears
from his phone. "Hello, Kaeya. The others	and	I have been wondering if you have
Harbor, he admired the bright shining lights	and	the merchants who were selling all sorts
a young lady with two brunette ponytails	and	shining reddish eyes. "Welcome to Wangsheng Funeral
hair with orange tips at the end	and	deep orange eyes. "A pleasure to meet

Fig. 4. Concordances 'and': Northern Region (34 of 1,400 lines)

[37] describes *a* as a determiner used when referring to someone or something for the first time or when people are unsure of who or what they are referring to. There are 1,050 concordance lines counted (Northern: 1,050). Through study, it was discovered that every advanced Malaysian upper primary school learner in Peninsular Malaysia mastered the correct and acceptable use of *a* as a determiner in their extended writing.

big and grey luxurious car driven by	a	drunken man. She was shocked and tried
car was totally fine because it was	a	big car. Also, the driver in that
Lisa still didn't wake up from fainting.	A	few minutes later, an ambulance and the
school vacations. I believe you should try	a	new sport, one that is aerobic in
exercise because it can help you maintain	a	healthy weight. When combined with a good
maintain a healthy weight. When combined with	a	good diet, aerobic exercise helps you lose
lose weight and keep it off. As	a	result, it can help you improve your
fitness, and strength. When you first begin	a	regular aerobic exercise, you may feel exhausted.
helps to minimise the chance of developing	a	variety of diseases. Obesity, heart disease, hypertension,
without purchasing anything; all you need is	a	yoga map. I hope you are able
caution with COVID-19. Take care and have	a	pleasant day.
involved in education. However, there is always	a	pot of gold at the end of
at the end of the rainbow. As	a	student myself, I believe online learning has
than that, unstable internet connections is also	a	major problem during an online class. Personally,
pandemic turned our high school experience into	a	far more stressful life than what it
our workspace that suits our like for	a	comfortable learning experience. On the other hand,
the other hand, online learning also has	a	lot of disadvantages. For example, it is
home just feels like we are having	a	holiday which interferes with our mindset to
the only way. This year has been	a	very hectic and problematic year for every
class via online learning. Online learning is	a	foreign way for students to learn. As
foreign way for students to learn. As	a	student, I get a lot of experience
to learn. As a student, I get	a	lot of experience during online learning. Online
online. Website such as Wikipedia is truly	a	lifesaver for me. I could get my
I could get my task done in	a	short period of time. Moreover, online learning
can be written using Microsoft Word provide	a	quick and easy way to detect errors.
attention from my class. Rainy weather is	a	major problem for me as it disrupts
the events will run smoothly. These are	a	few categories of awards that will be
subjects to make videos on conclusions or	a	summary of each topic in the textbook.
entertainment. This segment can feature students from	a	different club and asking questions on common
that he always uses to request for	a	taxi. He met them at the mall
excited after being unable to meet for	a	long time. Before the meeting happened, Adam
because of the whole pandemic. However, after	a	few weeks, the case seemed to drop
few weeks, the case seemed to drop	a	lot which led to the malls opening
at the mall. There were roller coasters,	a	haunted house and countless of amazing attractions.

Fig. 5. Concordances 'a': Northern Region (34 of 1,050 lines)

4. Discussions

The research question is: What are the differences and/or similarities in the use of English vocabulary in the extended writing produced by advanced learners in upper primary schools in the northern region of Malaysia?

Each difference and similarity of vocabulary (function & content) from the findings will now be discussed. As evidenced by [38] the significance of *the, to, is* that they play an essential function as articles, simple conjunctions, prepositions, and infinitives. Without the conjunctions the, to, and, a statement sounds awkward and may be grammatically wrong. Upper primary school learners prefer first-person pronouns rather than third-person ones in extended writing. It is in contrast to [39] because in his study, the majority of learners in upper primary school employed third-person pronouns for more extended writing. Despite this, *I* is frequently employed by secondary school students when they are required to provide their opinion and express their thoughts [40]. Possibly, learners in upper primary school are exposed to various types of essay writing and learning tools, resulting in the effective use of *I* in extended writing [41].

In reference to research question 1, it was reported that the learners in this study most recurrently used the infinitive 'to', conjunction 'and', and articles 'a' and 'the' that are fundamental for simple sentence formation in primary school. This aligns with findings obtained by [42] but on tertiary students who relied on using vocabulary including nouns and pronouns. Despite students producing a number of good vocabulary, students still tend to make mistakes in their writing because of carelessness. This study's conclusion is consistent with those of [43] who discovered that Malaysian upper primary school students did not employ English language to its full potential. As a result of their poor vocabulary skills, they were unable to compose a strong essay. The writing skills of the students were low intermediate.

Malaysian English primary school teachers struggled to come up with language and phrases to teach their students for guided and extended writing [44]. English lessons are taught in ELT classrooms by Malaysian English primary school teachers utilising terminology from the Get Smart Plus 4 (Year 4) and English Plus 1 textbook (Year 5). There was no additional vocabulary or teaching

of phrases in the English lessons. Students studying ESL in upper primary who are not exposed to sufficient phrases in the teaching and learning of English struggle to develop phraseology competence [45]. In the past 10 years, several researchers have started to investigate the vocabulary used by primary school students who are learning a language other than their own. An example of this is the vocabulary of Spanish that students who have learned English as a foreign language. For instance, the Spanish vocabulary used by primary school students was the subject of research by [46]. Even while this kind of research is still in its infancy, it could undoubtedly be expanded to include additional students that study ESL or something similar. According to [47], more research should be done on the vocabulary use of primary school students so that language educators can select and create the best pedagogical procedures, including materials for teaching and learning, for second language acquisition.

Numerous past studies on language learning, specifically those looking at EFL learners, have made known that learners at practically all skill levels struggle to use common verbs such as "make." Additionally, they suggest that activities using concordances can help learners become more conscious of the complexity of high-frequency verbs [48]. This study suggests that teachers must embed a detailed process of using corpora in language teaching merely for pedagogical goals as recommended by [49] in defining the "structure or language elements" for instruction. Teachers' role is to provide students with corpus-based activities for practice. Teachers will be able to identify how proficient a learner is through the corpus-based activities. [50] agrees that learner corpora should be applied in facilitating second language learning and error analysis to promote data-driven learning that will assist learners in becoming more aware of their native language and accelerate their language acquisition.

5. Conclusions

Based on the findings, it is clear that functional words use in vocabulary including functional and content words are identified. The advanced upper primary school learners in different areas in the northern regions are exposed to varied learning resources and materials, causing them to use different vocabulary and phrases for extended writing. The theme of the world of self, family, and friends is fundamental and essential, as most of the identified vocabulary and phrases revolve around it. Hence, it could be said that the use of salient vocabulary is highly influenced by the theme introduced in the school. Furthermore, it could be assumed that this theme is authentic to school learners as they are encouraged to write based on authentic experiences which makes the theme of the world of self, family and friends become the first and significant theme introduced by the English teachers since Year 1. Finally, the AMUPSLC could serve as a writing reference for Malaysian English language teachers and upper primary school pupils. The salient vocabulary and phrases from the findings can be used for extended writing activities.

References

[1] Friginal, Eric, ed. *Studies in Corpus-based Sociolinguistics*. London: Routledge, 2018. https://doi.org/10.4324/9781315527819

[2] Shauki, Baiti, and Manvender Kaur Sarjit Singh. "Developing a Corpus of Entrepreneurship Emails (COREnE) for Business Courses in Malaysian University Using Integrated Moves Approach." *Sains Humanika* 14, no. 1 (2022): 1-9. https://doi.org/10.11113/sh.v14n1.1885

[3] Foi, Liew Yon, and Teoh Hong Kean. "STEM education in Malaysia: An organisational development approach?." *International Journal of Advanced Research in Future Ready Learning and Education* 29, no. 1 (2022): 1-19.

[4] Spolsky, Ellen. *Contracts of fiction: Cognition, culture, community*. Oxford University Press, 2015.

[5] Borrego, M., J. E. Froyd, and T. S. Hall. "Ministry of Education Malaysia (2012)"Malaysia Education Blueprint 2013-2025,"." *CEE BOOK SERIES* (2012): 68.
[6] Aziz, Roslina Abdul, and Zuraidah Mohd Don. "Tagging L2 Writing: Learner Errors and the Performance of an Automated Part-of-Speech Tagger." *GEMA Online Journal of Language Studies* 19, no. 3 (2019). https://doi.org/10.17576/gema-2019-1903-09
[7] Zheng, Cui, and Tae-Ja Park. "An Analysis of Errors in English Writing Made by Chinese and Korean University Students." *Theory & Practice in Language Studies* 3, no. 8 (2013). https://doi.org/10.4304/tpls.3.8.1342-1351
[8] Malik, Mohd Azry Abdul, Nur Izzatulsyimah Madzuki, Nur Syahidah Shahnirul Hizam, Nuramanina Husna Shamsul Kamal, Nur Syaliza Hanim Che Yusof, Mohd Faiez Suhaimin, and Siti Nurani Zulkifli. "Teachers' Readiness and Practices in School-Based Assessment Implementation: Primary Education in Malaysia." *International Journal of Advanced Research in Future Ready Learning and Education* 23, no. 1 (2021): 1-9.
[9] Jamaludin, Aaishah Radziah, Wan'Atikah Wan Ibrisam Fikry, Siti Zhafirah Zainal, Fatin Shaqira Abdul Hadi, Nawal Shaharuddin, and Nurul Izzati Abd Rahman. "The effectiveness of academic advising on student performance." *International Journal of Advanced Research in Future Ready Learning and Education* 25, no. 1 (2021): 20-29.
[10] Zafar, Ameena. "Error analysis: a tool to improve English skills of undergraduate students." *Procedia-Social and Behavioral Sciences* 217 (2016): 697-705. https://doi.org/10.1016/j.sbspro.2016.02.122
[11] Sari, Syahar Nurmala, and Dyah Aminatun. "Students'perception On The Use Of English Movies To Improve Vocabulary Mastery." *Journal of English Language Teaching and Learning* 2, no. 1 (2021): 16-22. https://doi.org/10.33365/jeltl.v2i1.757
[12] Goundar, Prashneel Ravisan. "Vocabulary Learning Strategies (VLSs) Employed by Learners of English as a Foreign Language (EFL)." *English Language Teaching* 12, no. 5 (2019): 177-189. https://doi.org/10.5539/elt.v12n5p177
[13] Leech, Geoffrey N. *The pragmatics of politeness*. Oxford Studies in Sociolinguis, 2014. https://doi.org/10.1093/acprof:oso/9780195341386.001.0001
[14] Govindarajoo, Mallika Vasugi, Chow Chin Hui, and Siti Farhah A. Aziz. "Common Errors Made in English Writing By Malaysian Chinese Primary Year 6 ESL Learners At A Tuition Centre In Puchong, Malaysia." *Asian Journal of University Education* 18, no. 3 (2022): 674-691. https://doi.org/10.24191/ajue.v18i3.18954
[15] Siddek, Nur Amalina Jaafar, and Hanita Hanim Ismail. "Understanding Learners' Difficulties in Narrative Writing Among Malaysian Primary Learners." *Asian Journal of Research in Education and Social Sciences* 3, no. 2 (2021): 244-255.
[16] Yunus, Melor Md, and Chan Hua Chien. "The use of mind mapping strategy in Malaysian university English test (MUET) Writing." *Creative Education* 7, no. 04 (2016): 619. https://doi.org/10.4236/ce.2016.74064
[17] Miin, Wong Pei, Lee Yi Rou, and Melor Md Yunus. "Google Docs: Step by step sentence construction for primary school marginal passing rate pupils." *Creative Education* 10, no. 02 (2019): 237. https://doi.org/10.4236/ce.2019.102019
[18] Zhu, Yan, and Beilei Wang. "Investigating English language learners' beliefs about oral corrective feedback at Chinese universities: A large-scale survey." *Language awareness* 28, no. 2 (2019): 139-161. https://doi.org/10.1080/09658416.2019.1620755
[19] Prichard, Caleb, and Andrew Atkins. "Evaluating the vocabulary coping strategies of L2 readers: An eye tracking study." *TESOL Quarterly* 55, no. 2 (2021): 593-620. https://doi.org/10.1002/tesq.3005
[20] Ponomarenko, E. B., G. G. Slyshkin, E. A. Baranova, I. G. Anikeeva, and Y. V. Sausheva. "Linguistic and cultural analysis of the gender characteristics of british song slang." *XLinguae* 14, no. 2 (2021): 169-184. https://doi.org/10.18355/XL.2021.14.02.13
[21] Tognini-Bonelli, Elena. "Corpus linguistics at work." *Corpus Linguistics at Work* (2001): 1-236. https://doi.org/10.1075/scl.6
[22] Love, Robbie. *Overcoming challenges in corpus construction: The spoken British National Corpus 2014*. Routledge, 2020. https://doi.org/10.4324/9780429429811
[23] Malik, Muhammad Khan Abdul. *Interlanguage Error Analysis: an Appropriate and Effective Pedagogy for Efl Learners in the Arab World*. Xlibris Corporation, 2020.
[24] De Fina, Anna, and Alexandra Georgakopoulou, eds. *The Cambridge handbook of discourse studies*. Cambridge University Press, 2020. https://doi.org/10.1017/9781108348195
[25] Sarkar, Dipanjan. *Text analytics with python*. Vol. 2. New York, NY, USA:: Apress, 2016. https://doi.org/10.1007/978-1-4842-2388-8
[26] Karpenko-Seccombe, Tatyana. *Academic writing with corpora: A resource book for data-driven learning*. Routledge, 2020. https://doi.org/10.4324/9780429059926

[27] Han, Rui, and Yanlin Yin. "Application of web embedded system and machine learning in English corpus vocabulary recognition." *Microprocessors and microsystems* 80 (2021): 103634. https://doi.org/10.1016/j.micpro.2020.103634
[28] Lun, Wong Wei, Mazura Mastura Muhammad, Warid Mihat, Muhammad Syafiq Ya Shak, Mairas Abdul Rahman, and Prihantoro Prihantoro. "Vocabulary Index as a Sustainable Resource for Teaching Extended Writing in the Post-Pandemic Era." *World Journal of English Language* 13, no. 3 (2023): 181-181. https://doi.org/10.5430/wjel.v13n3p181
[29] Lun, Wong Wei, Mazura Mastura Muhammad, Muhamad Fadzllah Zaini, Ashrol Rahimy Damit, Carrine Teoh-Ong, Charanjit Kaur Swaran Singh, and Norhayati Yusoff. "Analysis of Covid-19 related Phrases Using Corpus-Based Tools: Dualisms Language & Technology." *Journal of Positive School Psychology* 6, no. 3 (2022): 5034-5044.
[30] Goyak, Flora, Mazura Mastura Muhammad, Farah Natchiar Mohd Khaja, Muhamad Fadzllah Zaini, and Ghada Mohammad. "Conversational mental verbs in english song lyrics: a corpus-driven analysis." *Asian Journal of University Education (AJUE)* 7, no. 1 (2021): 222-239. https://doi.org/10.24191/ajue.v17i1.12619
[31] Zaini, Muhamad Fadzllah, Anida Sarudin, Mazura Mastura Muhammad, and Siti Saniah Abu Bakar. "Representatif Leksikal Ukuran sebagai Metafora Linguistik berdasarkan Teks Klasik Melayu." *GEMA Online Journal of Language Studies* 20, no. 2 (2020). https://doi.org/10.17576/gema-2020-2002-10
[32] Zaini, Muhamad Fadzllah, Mazura Mastura Muhammad, Norliza Jamaluddin, Md Zahril Nizam Md Yusoff, Nordiana Hamzah, Noor Zuhidayah Muhd Zulkifli, Mohd Haniff Mohd Tahir, and Sharmini Pillai. "Protocol methodology for permission release in the construction of a written corpus." *MethodsX* 9 (2022): 101754. https://doi.org/10.1016/j.mex.2022.101754
[33] Low, Yu-Zane, Lay-Ki Soon, and Shageenderan Sapai. "A Neural Machine Translation Approach for Translating Malay Parliament Hansard to English Text." In *2020 International Conference on Asian Language Processing (IALP)*, pp. 316-320. IEEE, 2020. https://doi.org/10.1109/IALP51396.2020.9310470
[34] Hooi, Chee Mei, Helen Tan, Geok Imm Lee, and Sharon Sharmini Victor Danarajan. "Texts with Metadiscourse Features are More Engaging: A Fact or A Myth?." *3L, Language, Linguistics, Literature* 26, no. 4 (2020). https://doi.org/10.17576/3L-2020-2604-05
[35] Siti Syairah, Fakhruddin. "A corpus-informed study of discourse connectives in narrative essays of secondary school students/Siti Syairah Fakhruddin." PhD diss., University of Malaya, 2018.
[36] Morato, Jorge, Adrian Campillo, Sonia Sanchez-Cuadrado, and Ana Iglesias. "Influence of Term Familiarity in Readability of Spanish e-Government Web Information." In *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1*, pp. 905-915. Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-63128-4_68
[37] HarperCollins. Collins English Dictionary. (n.d.) https://www.collinsdictionary.com
[38] Saddhono, Kundharu, Muhammad Rohmadi, Budhi Setiawan, Raheni Suhita, Ani Rakhmawati, Sri Hastuti, and Islahuddin Islahuddin. "Corpus Linguistics Use in Vocabulary Teaching Principle and Technique Application: A Study of Indonesian Language for Foreign Speakers." *International Journal of Society, Culture & Language* 11, no. 1 (2023): 231-245. https://doi.org/10.22034/ijscl.2022.1971972.2823
[39] Struyk, Melinda W. "Exploring Writing Interventions for College Students." PhD diss., The University of Arizona, 2022.
[40] Pablo, Jim Christzer I., and Ronald Candy S. Lasaten. "Writing difficulties and quality of academic essays of senior high school students." *Asia Pacific Journal of Multidisciplinary Research* 6, no. 4 (2018): 46-57.
[41] Strobl, Carola, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Proske, and Christian Rapp. "Digital support for academic writing: A review of technologies and pedagogies." *Computers & education* 131 (2019): 33-48. https://doi.org/10.1016/j.compedu.2018.12.005
[42] Saddhono, Kundharu, Muhammad Rohmadi, Budhi Setiawan, Raheni Suhita, Ani Rakhmawati, Sri Hastuti, and Islahuddin Islahuddin. "Corpus Linguistics Use in Vocabulary Teaching Principle and Technique Application: A Study of Indonesian Language for Foreign Speakers." *International Journal of Society, Culture & Language* 11, no. 1 (2023): 231-245. https://doi.org/10.22034/ijscl.2022.1971972.2823
[43] Hasram, Syafiqah, M. Khalid M. Nasir, Maslawati Mohamad, Md Yusoff Daud, Mohd Jasmy Abd Rahman, and Wan Muna Ruzanna Wan Mohammad. "The effects of wordwall online games (Wow) on english language vocabulary learning among year 5 pupils." *Theory and Practice in Language Studies* 11, no. 9 (2021): 1059-1066. https://doi.org/10.17507/tpls.1109.11
[44] Lee, Icy, and Rui Eric Yuan. "Understanding L2 writing teacher expertise." *Journal of Second Language Writing* 52 (2021): 100755. https://doi.org/10.1016/j.jslw.2020.100755
[45] Meunier, Fanny, Julie Van de Vyver, Linda Bradley, and Sylvie Thouësny, eds. *CALL and complexity–short papers from EUROCALL 2019*. Research-publishing. net, 2019. https://doi.org/10.14705/rpnet.2019.38.9782490057542

[46] Barcroft, Joe, and Javier Muñoz-Basols, eds. *Spanish Vocabulary Learning in Meaning-oriented Instruction*. Routledge, 2021. https://doi.org/10.4324/9781315100364
[47] Binhomran, Kholoud, and Sultan Altalhab. "The Impact of Implementing Augmented Reality to Enhance the Vocabulary of Young EFL Learners." *JALT CALL Journal* 17, no. 1 (2021): 23-44. https://doi.org/10.29140/jaltcall.v17n1.304
[48] O'Keeffe, Anne, Brian Clancy, and Svenja Adolphs. *Introducing pragmatics in use*. Routledge, 2019. https://doi.org/10.4324/9780429342950
[49] Reppen, Randi. *Using corpora in the language classroom*. Cambridge University Press, 2010. https://doi.org/10.1017/9781139042789.003
[50] Huang, Li-Shih. "Taking stock of corpus-based instruction in teaching English as an international language." *RELC Journal* 49, no. 3 (2018): 381-401. https://doi.org/10.1177/0033688217698294