

---

## Mapping Vocabulary: A corpus-based comparison of Indonesian fourth-grade students' writing and English language textbooks

Fitria Ningsih

UIN Sayyid Ali Rahmatullah Tulungagung

e-mail: [ningsih.fitria.id@gmail.com](mailto:ningsih.fitria.id@gmail.com)

---

---

*Keywords:*

*corpus, frequency, writing, word-lists, vocabulary*

---

### ABSTRACT

This study compares the vocabulary used by Indonesian fourth-grade students in their free writing with that taught in their English language textbooks to assist language teachers in developing relevant learning materials for young learners. To achieve this, the researcher employed the corpus linguistic method and analyzed the writings of 369 students from 19 public elementary schools in Malang, East Java, Indonesia, along with five selected English textbooks. The University of Adelaide Text Analysis Tool program, AdTAT, was used to analyze the corpus. The program generated word lists consisting of the top 100 words most frequently used by students and the top 100 words used in English textbooks. The findings of the study revealed a 45% match between the top 100 words most frequently used by students and the top 100 words used in English textbooks. This suggests that there is a significant overlap between the vocabulary used in free writing and that taught in textbooks. The implication of this study is that language teachers need to recognize the basic words students acquire before learning new vocabulary to achieve successful learning of the target language. The study suggests that language teachers should incorporate more culturally relevant vocabulary into their teaching materials to better engage and motivate their students. The findings of this study can also assist language curriculum developers in designing more effective and culturally relevant teaching materials for young learners.

---

### 1. INTRODUCTION

The acquisition of vocabulary is a significant part of language learning and teaching, particularly for English as a Foreign Language (EFL) students who have multilingual backgrounds (Goundar, 2019; Sutrisna, 2021; Talbi Hassani, 2021). One of many challenges in teaching English to multilingual EFL students lies in the student's mastery of English vocabulary, where word meanings may not align with the contexts of the EFL learners' vocabulary learned earlier in other languages. Students need to have high absorption power to memorize the vocabulary. The more often it is applied, it will have a good impact on students including vocabulary, speaking and writing skills (Zamzami et al., 2023)

Furthermore, teaching and learning English as a foreign language for students who have a multilingual background should take this fact into account in addressing students' needs. Satullaeva & Kurbanbaeva (2020) stated that to effectively teach English as a second language, it is necessary to conduct a comparative analysis that allows students to recognize similarities and differences between the three languages used in the educational context. Learning multiple foreign languages should not be approached as separate endeavors but rather as an interconnected and interdependent study that involves comparing languages and building upon the linguistic knowledge and skills of the students.

Many non-native English teachers use language learning source materials that they consider appropriate for non-English background learning. The use of the mother tongue (L1) is then used to assist communication between teachers and students. This may have negative results, leading to inappropriate learning sources and meanings that are not relevant to the context (Hawanti, 2014). This commonly happens to children who are beginning to learn English. Furthermore, Chai et al., (2021) added in their study that it is crucial to have efficient methods and engaging vocabulary resources for young ESL learners in primary school.

Therefore, it's essential to identify appropriate techniques that meet their learning needs to achieve the satisfactory teaching of English vocabulary, for use by non-native speaking teachers of English. A study conducted by Dolmacı and Ertaf (2016) investigated that new Turkish language learners should learn academic vocabulary before they study at the undergraduate level. The Academic Turkish Word-list included 32% of the academic corpus and learning that word-list would help the students who want to learn Turkish to achieve better academic language skills, particularly reading comprehension. This approach also applies to EFL students, particularly to young learners who start to learn English as an additional language. Different strategies such as paying attention to word lists or completing various vocabulary tasks could be adopted by EFL students to learn vocabulary (Huong, 2018).

In Indonesia, for some students who are learning English as a foreign language, the use of the mother tongue exclusively becomes a barrier. The different backgrounds and cultural contexts between the two languages could become a major distraction for students when learning another language. An alternative method has been noted in studies that have been conducted using corpus approaches dealing with word frequency, and that have produced word lists for integrating into young learners' vocabulary. Firstly, Dang (2019) explores the current research on corpus-based word-lists for English language learners, including different types of word-lists, important issues such as a unit of counting and corpus construction, and how these word-lists can be applied in language learning and teaching. The research identifies aspects of wordlist study that demand more exploration, particularly mid-frequency, and subject-specific vocabulary. It also underlines the necessity of combining objective and subjective criteria during the construction of word-lists.

Secondly, a corpus-based study was conducted by Tiliakou & Frantz (2021) who analyzed writing samples from 12 and 13-year-old Greek learners of English. Their objective was to evaluate the level and quantity of vocabulary used in the writing samples and to determine whether it corresponds to the expected proficiency level based on the Common European Framework levels A1-A2. The corpus was analyzed using keyword analysis and the English Vocabulary Profile Online tool. The results show that most of the content words used belonged to the A1 level, while a smaller amount was of the A2 level and a few words to B1 and B2 levels. The study found that higher-level words were correlated with higher graded writings and higher proficiency of the learners.

Thirdly, a corpus study conducted by Berkling (2016) describes the H1 Corpus, which is a collection of weekly writing samples from 2nd and 3rd graders aged 7-11, collected over 3 months in a classroom setting by the teacher. The corpus contains written transcriptions of texts that include both correct spelling and spelling errors, as well as scanned-in versions. It's accessible for research purposes through the Linguistic Data Consortium, and scholars are encouraged to add more annotations and enhancements to the corpus before sharing it with the public domain via LDC.

Furthermore, some experts have suggested solutions for how to teach English vocabulary as influenced by the first language (L1) or the second language (L2). One of the practical ways to achieve this is through word listing, produced through a corpus (Dang, 2019). Overall, the aims are to prepare vocabulary for the materials as suitable for the students' levels and not too difficult to use in classrooms.

However, from these findings, when focusing on early childhood learning, the target levels of the students may be different. Therefore, in assessing the research gaps in this area, this study has created a corpus of the most frequent words used in Indonesian fourth-grade students' free writing, when analyzed and compared to the vocabulary taught in English language textbooks. Grade 4 is an appropriate time for English vocabulary learning in Indonesia, in line with the national curriculum standards for English language education (P et al., 2018). At this stage, students have a basic understanding of English phonetics, vocabulary, and grammar elements. According to Rahmat & Mohandas (2020), vocabulary acquisition among young learners is critical for language learning and academic success, and early exposure to English vocabulary improves their proficiency in the language. Many language education experts also recommend prioritizing vocabulary acquisition in the early stages of language learning. The inclusion of English language education in the national curriculum for primary schools in Indonesia aims to equip students with the necessary language skills to communicate effectively in a globalized world (Ministry of Education and Culture of the Republic of Indonesia, 2013). As grade 4 students are in a crucial stage of language development, they require adequate support and resources to learn English vocabulary effectively. This study creates a corpus of the most frequent words used by grade 4 students in their free writing and compares it with the vocabulary taught in English language textbooks. The study aims to provide valuable insights into the vocabulary acquisition process among young learners, assisting language teachers in developing more effective and culturally relevant learning materials.

## **2. METHOD**

This study employed corpus in linguistics as the methodology to gather the data and the analyses. According to Waskita (2017), corpus linguistics is a method to investigate any kind of questions dealing with linguistic problems which could provide new insight into language use. Cushing's (2017) work discusses the main objective of corpus linguistic research, which is the exploration of language usage, as well as the sharing of research findings that pertain to the cumulative effects of natural language and the examination of frequency and distributional information.

This study adopted mixed quantitative and qualitative methods. Mixed methods research represents that there are several approaches to combining qualitative and quantitative research methods. These include using qualitative methods to identify constructs or hypotheses that can be studied in quantitative research, using quantitative studies to uncover patterns that can be further investigated in qualitative research, and converting qualitative data into quantitative data through quantification (Kaplan, 2014). Quantitative data will be collected from the list of most frequent words used by Indonesian students and then analyzed using an application, AdTAT, the Adelaide Text Analysis Tool, a concordance tool provided by the University of Adelaide. The most frequent words used were categorized into a top 100 words list. The word lists are presented in tables which are completed with the numbers of frequency and the meaning. For the qualitative data, the study presented an analysis of the quantitative

data from the most frequent words used in the students' writing and for comparison, the English textbooks.

Moreover, the sampling method used in this study is cluster sampling. Malang Region, East Java, Indonesia, was selected as the study area due to its diverse socio-economic backgrounds and which are known for its diverse student populations and adherence to the national English language curriculum standards. From the list of public elementary schools in the Malang Region, 19 schools were randomly selected and invited to participate in the study. Once the schools agreed to participate, the researchers obtained written consent from the school administrators. Then, a school code was assigned to each participating school, and the students were instructed to write a short composition in English about a given topic. The samples of students' writing were collected by the researchers, who then recorded the datasets using the school codes to maintain anonymity and confidentiality.

Furthermore, the sample size of 369 students was determined based on the power analysis calculation, which took into consideration the total number of students in the participating schools. The students' age range was between 9 to 10 years old, which corresponds to the fourth year of elementary school in Indonesia. Of these students, 165 were boys and 214 were girls. The data collected from the writing samples were then analyzed and compared to the vocabulary taught in English language textbooks. To collect the data, the researchers provided each student with a writing prompt and asked them to write freely for 20-30 minutes. As it was free writing, students were free to decide the topic on which they wanted to write. However, in some schools, the teachers assisted the students to create writing dealing with their real-life contexts, their holidays, and stories about their past experiences, their hobbies, or their future ideas.

### **3. RESULTS AND DISCUSSION**

These findings are then divided into points relating to the word frequency from the Indonesian students' writing and the English textbooks used in the schools and the analysis of the vocabulary in terms of content and cultural views used in the textbooks when compared to the students' writing.

#### **3.1 Results**

The first research question asked about the most frequent Indonesian words used by Indonesian year 4 students in writing and the most frequent words taught in their English textbooks. As 19 schools participated in this study which involved 369 students' writing as the samples for the corpus, it collected 33916 words. From the five selected English textbooks used as the sample for the analysis in this study, the researcher gained 8270 words, with the vocabulary collected from the main sections: introduction, dialogues, reading, and the vocabulary list. Thus, the findings produced the top 100 most frequent words used by the students in writing Indonesian texts and the top 100 most frequently used vocabulary taught from the English language textbooks.

##### **3.1.1 Top 100 Most Frequent Words from the Indonesian Students' Writing and the English Textbooks**

After using AdTAT software to collect words frequency lists from the samples of the student's writing and the English textbooks, below are the top 100 words used most frequently in students' writing (See Table 1)

**Table 1.** Top 100 words frequency from students' writing

No	Words		Frequencies
	Indonesian	English	
1	saya	I	1857
2	dan	and	1531
3	aku	I	1241
4	di	at/in/on	1031
5	ke	to	799
6	sana	there	562
7	yang	that	487
8	lalu	then	425
9	itu	that/those	360
10	bermain	play	315
11	dengan	with	313
12	cita-cita	idea	307
13	rumah	house	306
14	setelah	after	285
15	teman	friend	281
16	sangat	very	276
17	bersama	together with	272
18	keluarga	family	261
19	pulang	go home	252
20	ada	there is/are	249
21	sekali	very/once	236
22	liburan	holiday	231
23	hari	day	228
24	makan	eat	228
25	pergi	go	216
26	pantai	beach	206
27	senang	happy	193
28	melihat	see	191
29	untuk	for/to	187
30	orang	people	184
31	ayah	father	182
32	ingin	want	180
33	kami	we/us/our	178
34	juga	also	171
35	menjadi	be	161
36	Ibu	mother	152
37	pada	at/to also	147
38	nenek	grandmother	146
39	berenang	swim	144
40	karena	because	141
41	bisa	can	141
42	sudah	already	140
43	sekolah	school	138
44	sampai	arrive/till	132
45	saat	when	131
46	tidak	no/not	129
47	sama	same	129
48	ku	my	124
49	jam	hour/clock	122
50	kakak	big brother/ sister	122

No	Words		Frequencies
	Indonesia	English	
51	banyak	many/ much	114
52	berlibur	go holiday	111
53	anak	children	107
54	naik	ride	104
55	adik	little brother/ sister	104
56	alun-alun	town square	100
57	layang	kite	99
58	diajak	asked	97
59	indah	beautiful	92
60	mandi	take a bath	91
61	jalan	road	90
62	dari	from	90
63	terus	then/next	85
64	langsung	straight	82
65	baju	clothes	80
66	pun	even	78
67	kakek	grand- father	78
68	ikan	fish	77
69	kita	we/us/our	77
70	bola	ball	75
71	membeli	buy	75
72	besar	big	75
73	lagi	again	72
74	dokter	doctor	70
75	perjalanan	trip	69
76	tidur	sleep	68
77	harus	must	68
78	tapi	but	63
79	mobil	car	62
80	beli	buy	62
81	tiba	arrive	62
82	waktu	time	61
83	nama	name	60
84	akan	will	60
85	belajar	study	59
86	semua	all	59
87	Malang	Malang	59
88	sakit	sick	58
89	dia	she/he	58
90	habis	finished	57
91	oleh	by	57
92	minggu	Sunday	56
93	foto	photo	56
94	kalau	if	55
95	pemandangan	view/ scenery	54
96	saudara	siblings	53
97	kemudian	then/next/	53
98	berangkat	depart/ leave	53
99	guru	teacher	53
100	lain	other	53

From the results of the top 100 words frequency from the students' writing, most of the students expressed their writing using the word *I* which became the top 1 and 3 *saya* and *aku*

respectively. The words *saya* and *aku* in Indonesia have the same meaning as *I* in English. These two words are used widely by Indonesian in daily conversation. However, these two words have different uses. The word *saya* is normally used in formal ways such as in lecturing, seminars, speech, or in other formal spoken and written contexts. The word *aku* is less formal compared to the word *saya*. It is then followed using the preposition *in/at/on* and some adverbs, which precede the use of verbs and nouns.

The first verb in the top 10 is the word *play*. The students use the word *play* hand in hand with the word *with* in the following list. It could be related to the word *friend* in the top 15 words in the list. Thus, most students said that they prefer to use *play with a friend* in their writing. Moreover, it could be linked to the word *family* in the top 18, as it can be assumed that the students like to play with family. This is followed by the word *holiday* in the top 22. Therefore, most of the students like to play with their friends and family during the holidays. About the word *holiday*, the word *beach* followed in the list of the top 25 of the most frequent words used. It also indicated that most of the students like having a holiday, particularly at the beach, with *swim* in the top 39, which can indicate that they were at the beach where they *play* and *swim*. This could be related to the events that happened when the data was gathered. It was when students had long holidays during the school break, so the findings showed some words related to a holiday.

The students also use the word *eat* which appears 228 times. However, there is no word *drink* in the top 100 most frequent words where the word *eat* is correlated to *drink*. Students used verbs such as *go*, *see*, and *ride* in the top 55 words. The other verbs in the top 100 words used frequently are *take a bath*, *buy*, and *sleep*.

The use of nouns started with the word *cita-cita* or *idea* in the future in the top 12 of the most frequent words list. In the top 100 words, the majority of students had an idea of a *doctor* which come up 70 times on the top 74 lists. Following this, students wrote vocabulary dealing with *house* in the top 13. In line with the people in the house, students mostly talked about family members. As shown in Table 1, the word *father* appears 182 times, more often than *mother* at 152 times, 30 less than the word *father*. This is followed by the word *grandmother* which appears 146 times, *big brother/sisters* 122 times, *grandfather* 78, times, and is siblings 53 times. Most of the students are engaged with their family members in the house.

**Table 2.** Top 100 words frequency from English textbooks

No.	Words	Frequencies
1	is	508
2	the	315
3	a	228
4	I	208
5	my	183
6	you	143
7	are	135
8	it	130
9	in	119
10	and	100
11	what	97
12	to	88
13	do	83
14	there	82
15	am	67
16	of	67

No.	Words	Frequencies
17	he	65
18	she	63
19	your	60
20	name	58
21	We	57
22	like	51
23	have	47
24	on	44
25	this	44
26	yes	42
27	can	40
28	school	39
29	at	37
30	good	34
31	me	34
32	room	33
33	that	33
34	her	32
35	they	32
36	no	30
37	today	29
38	for	29
39	table	29
40	father	28
41	day	27
42	go	27
43	how	26
44	student	26
45	teacher	25
46	number	25
47	month	25
48	bag	24
49	mother	24
50	one	24
51	classroom	24
52	morning	23
53	clean	22
54	book	22
55	his	21
56	also	21
57	not	21
58	want	21
59	new	20
60	where	20
61	sister	20
62	play	20
63	twenty	20
64	so	20
65	with	20
66	friend	19
67	very	19
68	our	19
69	class	19
70	see	18



No.	Words	Frequencies
71	nice	17
72	house	17
73	theme	17
74	chair	17
75	first	17
76	eat	17
77	don't	17
78	thank	16
79	every	16
80	Sunday	16
82	English	16
83	may	16
84	living	16
85	favourite	16
86	home	15
87	family	15
88	let's	15
89	many	15
90	picture	15
91	Doni	15
92	Hi	15
93	was	15
94	after	15
95	hello	14
96	wear	14
97	wall	14
98	five	14
99	has	13
100	beautiful	13

From the top 100 most frequent words used in English textbooks, students are taught mostly to use the word *be* which appears 508 times from a total of 8270 words. Some most frequent words that appear are *my, you, it, she, he, and your* which represent the use of pronouns. Until the top 20 of the most frequent words list, the textbooks use the word *name* as a noun that appears for the first time. This is reversed in the students' writing, in which the use of the noun comes first in the top 12.

Some of the verbs are taught to the students such as *like, have, can, and go* appear in the top of 50 of the word lists. The use of word *play* appears in the top 62 of the 100 most frequent words, which is different from the students' writing where it comes in the top 10 and first in the list of verbs used in their writing. However, like the order of students' writing, the word *play* is followed by the word *with* in the top 65, which is used hand in hand with the word *friend* in the top 66. This indicated that students *like to play with a friend* in the same word order in the results of the most frequent words in the students' writing.

Some nouns are often found between the top 37 and the top 55. Some words taught by the textbooks are *table, father, day, student, teacher, number, month, bag, mother, classroom, morning, and book*. Similar to the students' writing, the textbooks teach students to use the word *father* more often than *mother* as the word *father* comes in the top 40 and the word *mother* in the top 49. The words *student* and *teacher* are respectively in the 100 most frequent words of the English textbooks. It means that both student and teacher are equally used, mainly in the noun

list for the classroom. From the top 100 most frequent words in English textbooks, students are mostly taught the use of prepositions, pronouns, adverbs, nouns, and some verbs and adjectives.

### 3.1.2 Words used in top 100 most frequent words from both the Indonesian students' writing and the English textbook

From the most frequent words used both in students' writing and the English textbooks, the researcher classified and ranked 45 words, from the most frequent to the less frequent as shown in Table 3. The words are ranked by the sum of the number of percentages from both corpora. Firstly, the researcher compared the top 100 words list of the students' writing with the results of the top 100 most frequent words in English textbooks. Secondly, the researcher compared the top 100 words list of the English textbooks to the result of the TOP 100 most frequent words of the student's writing. It can be seen the similar words used ranked from the most frequent to the less frequent respectively.

From Table 3, the word *I* appears at the top of the lists. The use of the pronoun *I* is found in 6.19% of the overall students' writing. Moreover, the word *I* in the students' writing occurs 2098 times when this is a combination using the words *saya* and *aku*. In English textbooks, the word *I* occurs, just 208 times, which is 2.52% of the overall text in the textbooks and less used than the students' writing. However, the word *I* is the number 1 word of similar words used in both corpora.

The use of *to be (is)* becomes the second in the top 2 of the most similar words. However, the percentages of the students' writing are different. As seen in Table 3, the use of the word *is* or *be* in the students' writing reached 0.49% of the total words, while in the English textbooks reached 6.14% which is 12 times more frequent than in the students' writing. Some equally used words both in students' writing and the English textbooks are *can* which reached 0.43% and 0.48%, *school* which reached 0.42% and 0.47%, *Sunday* which reached 0.17% and 0.19%, and lastly, *picture* which reached 0.17% and 0.18% from the student's writing and the English textbooks respectively.

**Table 3.** The Rank of 45 most frequent words used both in the students' writing and the English textbooks

Words	Students' Writing		English Textbooks		Average	Combined Rank
	Frequencies	Percentage of 33916 Words	Frequencies	Percentage of 8270 Words		
I	2098	6.35%	208	2.52%	4.43%	1
is	161	0.49%	508	6.14%	3.32%	2
and	1531	4.64%	100	1.21%	2.92%	3
in	1178	3.57%	119	1.44%	2.50%	4
on	1178	3.57%	44	0.53%	2.05%	5
at	1148	3.48%	37	0.45%	1.96%	6
to	799	2.42%	88	1.06%	1.74%	7
that	847	2.57%	33	0.40%	1.48%	8
there	562	1.70%	82	0.99%	1.35%	9
my	124	0.38%	183	2.21%	1.29%	10
are	161	0.49%	135	1.63%	1.06%	11
we	255	0.77%	57	0.69%	0.73%	12
am	161	0.49%	67	0.81%	0.65%	13
play	315	0.95%	20	0.24%	0.60%	14
with	313	0.95%	20	0.24%	0.59%	15

Words	Students' Writing		English Textbooks		Average	Combined Rank
	Frequencies	Percentage of 33916 Words	Frequencies	Percentage of 8270 Words		
house	306	0.93%	17	0.21%	0.57%	16
friend	281	0.85%	19	0.23%	0.54%	17
very	276	0.84%	19	0.23%	0.53%	18
after	285	0.86%	15	0.18%	0.52%	19
day	228	0.69%	27	0.33%	0.51%	20
our	255	0.77%	19	0.23%	0.50%	21
go	216	0.65%	27	0.33%	0.49%	22
family	261	0.79%	15	0.18%	0.49%	23
he	58	0.18%	65	0.79%	0.48%	24
she	58	0.18%	63	0.76%	0.47%	25
for	187	0.57%	29	0.35%	0.46%	26
can	141	0.43%	40	0.48%	0.46%	27
eat	228	0.69%	17	0.21%	0.45%	28
father	182	0.55%	28	0.34%	0.44%	29
school	138	0.42%	39	0.47%	0.44%	30
name	60	0.18%	58	0.70%	0.44%	31
want	180	0.55%	21	0.25%	0.40%	32
see	191	0.58%	18	0.22%	0.40%	33
also	171	0.52%	21	0.25%	0.39%	34
no	129	0.39%	30	0.36%	0.38%	35
mother	152	0.46%	24	0.29%	0.38%	36
not	129	0.39%	21	0.25%	0.32%	37
big/little brother sister	104	0.31%	20	0.24%	0.28%	38
many	114	0.35%	15	0.18%	0.26%	39
be	161	0.49%		0.00%	0.24%	40
teacher	53	0.16%	25	0.30%	0.23%	41
beautiful	92	0.28%	12	0.15%	0.21%	42
Sunday	56	0.17%	16	0.19%	0.18%	44
picture	56	0.17%	15	0.18%	0.18%	45

Furthermore, in the findings, the preposition *di* in Indonesian has multiple meanings in English such as *at, in, and on* and this makes a difference when used in the Indonesian and the English language. The researcher then collated the multiple meanings in the list as shown in Table 3. In line with this, the use of the pronouns *we/us/our* which are differently used in Indonesian are *kami* and *kita, she, and he* which have one meaning in Indonesian as *dia*, as well as *brother* and *sister* could mean *big and little brother and sister*.

Both the students' writing and the English textbooks used most similar words in pronouns, prepositions, and adverbs as Table 3. shows from 1 to 13. The use of verbs first appears in the top 14 and nouns in the top 16. The combination of nouns and verbs are found in the top 17 as seen in Table 3.

### 3.2 Discussion

One of the aims of this study has been to collect the most frequent words used by Indonesian teachers when compared to the vocabulary taught in English language textbooks. Nation (2016) agreed with this goal, observing that the advantage of learning the most frequent words in use, will be that learners will have the most significant opportunities to enrich their knowledge of the words and produce it by themselves. Moreover, the target of this corpus study

has addressed the needs of the participants, that is, year 4 elementary students. As Nation (2016) suggested, When creating vocabulary lists for young children, it's crucial to base them on a corpus that accurately represents the language that children encounter in their daily lives. This aligns with the findings of the corpus study, which used Bahasa Indonesia, the student's first language, to collect the data sample. Lastly, Nation (2016) also suggested that when creating a corpus for young learners, it's beneficial to ensure that the corpus includes a similar proportion of different text types as the types of language that children encounter in their daily lives. This also meets the criterion of the data collection, in which the researcher used students' free writing which included text which recounted the students' experiences and the narrative texts of the stories that students had developed.

Concerning the vocabulary used by the students, O'Brien et al. (2014) stated that children's vocabulary knowledge varies greatly when they start school, and the extent of their early vocabulary knowledge has important consequences for their language development, both in the short term and in the long term. This language development could be linked to the language that students want to learn which, in this study, could be both the Indonesian and English languages. Moreover, the importance of word frequency revealed in a corpus of first-language words can improve an understanding of which words are used with which other words (Cameron, 2011). Thus, this study highlighted the vocabulary that students have already known as the basis for teaching vocabulary that is appropriate to the textbooks.

Some findings from this study show some compatibility with some previous theories and studies. Firstly, the top 100 most frequent words from the student's writing and the English textbooks both shared a similar use of the pronoun I, which is in line with (Liswahyuningsih & Bal, 2015) who stated that the use of the word I in the L1 language, may also occur when learning a foreign language. In telling narratives, children still need to learn how to create thematic structure through language, while still developing pronouns and determiners, In the top 100 most frequent words, the use of the word di in Indonesian could be translated into multiple meanings in the English language, such as at/in/on. The word I in English could be used as *saya* or *aku* in the Indonesian language. This is in line with the theory that different languages have different ways of providing meaning (Regier et al., 2015).

Furthermore, many theories claim that the correlation between L1 acquisition and L2 acquisition influences the vocabulary of language learning the schema is usually constructed throughout childhood, within the first language culture (Chavez, 2016; Jordan, 2016). This means that when a foreign language is learned, it will be mapped based on the first language words that have already been built up. Moreover, a concept of L2 acquisitions proposed by (Kim & Harris, 2014) stated that children tend to learn indirectly and to learn information from everything around them. This is very important for the teacher to know what students have learned, particularly the vocabulary that they have already acquired. The wordlist of the top 100 most frequently used words by Indonesian students in free writing could be used by teachers as guidance in teaching English as a foreign language.

It can be seen from the findings of the study that this could be used as a comparison with the initial studies which also had the same interest in language acquisition. In the findings, the word list produced from the English textbooks found that the use *to be* or *is*, occurred most frequently. This is in line with the Panggabean (2015) study which revealed that Indonesian-speaking learners of English tend to use *is*, *am*, and *are* regardless of different tenses and types of sentences they build up, and which there are commonly found to be mistaken in using and distinguishing them from other auxiliary verbs.

The study is like the findings revealed by which investigated L1 and L2 learning in L2 languages such as Germanic, Japanese, Turkish, and Indonesian, through writing, and found that there were similarities in using lexical and cohesive patterns. Regarding the correlation of the L1 and L2 acquisitions, the findings determined a 45% comparison between the top 100 most frequent words of the student's writing and the English textbooks. This finding also revealed that the two different languages share similar uses of the basic concept of vocabulary. This statement is related to the investigation conducted by Guo and Huang (2020), which utilized bilingual and learner analysis techniques to provide evidence of the influence of a learner's first language on their advanced writing in a second language. In particular, the research showed a connection between two distinct language studies within the corpus and the impact it has on the process of acquiring a new language.

It has been stated before that another aim of this study was to investigate the vocabulary taught in English textbooks compared to the words that students have already acquired, through their writing. Kim & Harris (2014) stated that the positive use of L1 could lead to positive foreign language learning when vocabulary learning begins around a topic from the learners rather than from the textbook. This means that the teacher can ask students which words they already know and which words they want to learn in a foreign language. This is in line with the methodology of this study which looks at the words L1 learners know from their writing, as a base for creating word-lists as guidance for learning other vocabularies.

Lastly, according to Panggabean (2015), in his study of EFL teaching and learning, for some elementary students, the materials and vocabularies need to be introduced from what they see every day in the classroom and their neighborhoods, from their daily activities to contextualize the materials within the students' context. This could be related to the results of the most frequent words used by the students when dealing with the classroom situations, such as *teacher, students, school, class, bag, and books*.

From the findings, we could see the classification of the topics distributed in the five selected textbooks. The vocabularies taught from the topics are varied and some of the textbooks shared the same topics such as greetings, number, calendar, family, schools, and food. Butler (2022) claimed that when teaching the vocabulary of a topic, it is better to begin with basic-level items, for example, food, such as pizza or rocket. This is because the vocabulary of the course books is predictable, particularly for young learners.

#### 4. CONCLUSION

The study analyzed the top 100 most frequently used words in the free writing of Indonesian year 4 students and English language textbooks. The study found that the lists of the top 100 most frequently used words in the two different language systems had a 45% similarity. The use of the pronoun "I" was the most frequent occurrence in both corpora, with the use of the word "*saya*" and "*aku*" in Indonesian students' writing.

The study has significant implications for the Indonesian education system. The top 100 most frequent words used by Indonesian students could be developed into teaching materials such as posters, which could be used in classrooms throughout Indonesia, particularly in Malang. Additionally, the corpus could be published for learners who are interested in learning new words in Bahasa Indonesia, providing valuable resources to help them improve their language skills.

However, the study's limitations suggest that the findings may not be widely applicable due to the diverse cultural and linguistic backgrounds in Indonesia, and further research is

needed to confirm the findings in other regions. Nevertheless, the study provides a foundation for future research that could improve language education in Indonesia. The use of corpus in linguistics can guide analyzing students' writing and textbooks, which can be beneficial for language education.

## REFERENCES

- Berkling, K. (2016). Corpus for children's writing with enhanced output for specific spelling patterns (2nd and 3rd Grade). *International Conference on Language Resources and Evaluation*.
- Butler, Y. G. (2022). Teaching vocabulary to young second- or foreign-language learners. *Language Teaching for Young Learners*, 4–33. <https://doi.org/10.1075/ltyl.00003.but>
- Chai, Z. F., Rosman, N. B., & Basri, M. F. Bin. (2021). Using word recognition instruction and visual aids to enhance Malaysian Primary ESL Learners' Vocabulary Achievement. *JETLE (Journal of English Language Teaching and Learning)*, 3(1), 28–39. <https://doi.org/10.18860/jetle.v3i1.13392>
- Chavez, M. (2016). The first language in the foreign language classroom: teacher model and student language use – an exploratory study. *Classroom Discourse*, 7(2), 131–163. <https://doi.org/10.1080/19463014.2016.1149499>
- Cushing, S. T. (2017). Corpus linguistics in language testing research. *Language Testing*, 34(4), 441–449. <https://doi.org/10.1177/0265532217713044>
- Dang, T. N. Y. (2019). Corpus-Based Word Lists in Second Language Vocabulary Research, Learning, and Teaching. In *The Routledge Handbook of Vocabulary Studies* (pp. 288–303). Routledge. <https://doi.org/10.4324/9780429291586-19>
- Dolmacı, M. & Ertas, A. (2016). Developing a Textbook-based Academic Turkish Wordlist, *In Procedia - Social and Behavioral Sciences*, Volume 232, 2016, Pages 821-827, ISSN 1877-0428, Retrieved from (<http://www.sciencedirect.com/science/article/pii/S1877042816313441> )
- Goundar, P. R. (2019). Vocabulary Learning Strategies (VLSs) Employed by Learners of English as a Foreign Language (EFL). *English Language Teaching*, 12(5), 177. <https://doi.org/10.5539/elt.v12n5p177>
- Guo, X., & Huang, L.-S. (2020). Are L1 and L2 strategies transferable? An exploration of the L1 and L2 writing strategies of Chinese graduate students. *The Language Learning Journal*, 48(6), 715–737. <https://doi.org/10.1080/09571736.2018.1435710>
- Hawanti, S. (2014). Implementing Indonesia's English language teaching policy in primary schools: The role of teachers' knowledge and beliefs. *International Journal of Pedagogies & Learning*, 9(2), 162-170. doi:10.5172/ijpl.2014.9.2.162
- Huong, L. P. H. (2018). A Survey Study on Academic Vocabulary Learning Strategies by EFL University Students. *Journal of Language Teaching and Research*, 9(5), 1009. <https://doi.org/10.17507/jltr.0905.15>
- Jordan, S. (2016). *L1 vs. L2 Acquisition: Reconciling the Differences between Childhood Language Development and Classroom Learning*.
- Kaplan, S. (2014). Mixing Quantitative and Qualitative Research. *Political Methods: Qualitative & Multiple Methods EJournal*.
- Kim, S., & Harris, P. L. (2014). Children prefer to learn from mind-readers. *British Journal of Developmental Psychology*, 32(4), 375–387. <https://doi.org/10.1111/bjdp.12044>

- Liswahyuningsih, N. L. G., & Bal, I. P. (2015). Semantics Analysis of Pronoun I and You in Balinese Language. *International Journal of Research in Social Sciences*, 5, 518–531.
- Nation, I. (2016). Making and using word lists for language learning and testing. Retrieved from <https://ebookcentral.proquest.com/lib/adelaide/detail.action?docID=4673371>
- O'Brien, L., M., et al. (2014) Examining Differential Effects of a Family Literacy Program on Language and Literacy Growth of English Language Learners with Varying Vocabularies. *Journal of Literacy Research* Vol 46, Issue 3, pp. 383 - 415 First Published October 7, 2014. Retrieved from <http://journals.sagepub.com.proxy.library.adelaide.edu.au/doi/abs/10.1177/1086296X14552180#articleCitationDownloadContainer>
- P, V. U., Nurnaningsih, A. A., & I, A. P. (2018). Learning English Vocabulary through Online Games: Case Study of Students In 4th Grade of State Elementary School (SDN) Jombor 01, Sukoharjo, Central Java, Indonesia. *International Journal of Multicultural and Multireligious Understanding*, 5(4), 470. <https://doi.org/10.18415/ijmmu.v5i4.433>
- Panggabean, H. (2015). Problematic Approach to English Learning and Teaching: A case study in Indonesia. *English Language Teaching*; Vol. 8, No. 3; 2015 ISSN 1916-4742 E-ISSN 1916-4750 Published by Canadian Center of Science and Education. Retrieved from <http://www.ccsenet.org/journal/index.php/elt/article/view/45413/24616#>
- Rahmat, N. H., & Mohandas, E. S. (2020). Vocabulary acquisition among esl learners: A look at strategies and barriers. *European Journal of English Language Teaching*, 6(2). <https://doi.org/10.46827/ejel.v6i2.3431>
- Regier, T., Kemp, C., & Kay, P. (2015). Word Meanings across Languages Support Efficient Communication. In *The Handbook of Language Emergence* (pp. 237–263). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118346136.ch11>
- Satullaeva, N., & Kurbanbaeva, G. (2020). Teaching English As A Second Foreign Language. *The American Journal of Social Science and Education Innovations*, 02(08), 351–357. <https://doi.org/10.37547/tajssei/Volume02Issue08-59>
- Sutrisna, G. (2021). Vocabulary acquisition in EFL: A literature review of innovative vocabulary teaching strategies. *Yavana Bhasha: Journal of English Language Education*, 4(1), 8. <https://doi.org/10.25078/yb.v4i1.2215>
- Talbi Hassani, A. (2021). Technology as a source of language input and practice and its influence on the vocabulary development of university EFL students. *International Journal of Recent Contributions from Engineering, Science & IT (IJES)*, 9(4), 19–30. <https://doi.org/10.3991/ijes.v9i4.24715>
- Tiliakou, C. S., & Frantzi, K. T. (2021). *Investigation of Vocabulary in a Corpus of Written Production of Greek Learners of English*.
- Waskita, D. (2017). Corpus linguistics: Method, theory, and practice. *Jurnal Socioteknologi*, 16(1), 145–147. <https://doi.org/10.5614/sostek.itbj.2017.16.1.12>
- Zamzami, M. I., Krenata, N. A. C., & Rohmana, W. I. M. (2023). The use of poetry in English learning for Islamic junior high school students | JELITA. 4(1). <https://jurnal.stkipmb.ac.id/index.php/jelita/article/view/108>