

## Vocabulary in the BIPA Book "Sahabatku Indonesia": A Corpus-Based Study

Dadang Rhubido<sup>1</sup>, Kisyani Laksono<sup>2</sup>, Syamsul Sodiq<sup>3</sup>

<sup>1,2,3</sup>Universitas Negeri Surabaya

<sup>1</sup>dadangrhubido@unesa.ac.id; <sup>2</sup>kisyani@unesa.ac.id; <sup>3</sup>syamsulsodiq@unesa.ac.id

### Abstract

This research aims to describe a number of vocabulary words contained in the BIPA book "*Sahabatku Indonesia*" levels BIPA 1 to BIPA 7 published by Kemendikbudristek. This vocabulary list can later be used as a basis for compiling a BIPA dictionary or can be used as a learning medium for vocabulary that BIPA students must master. This research uses mixed methods, namely combining a quantitative approach and a qualitative approach by utilizing corpus linguistic research methods. The research data is in the form of vocabulary contained in the book BIPA Sahabatku Indonesia levels 1 to level 7 published by the Ministry of Education and Culture. Based on the results of research and analysis, it is known that (1) not all book levels experience an increase in vocabulary. The number of tokens and the number of dictions at each level do not show consistency; (2) the word class that appears most frequently is a noun; (3) the ten highest frequency vocabularies are conjunctions, prepositions, and pronouns; (4) the highest frequency vocabulary in the *Sahabatku Indonesia* book is similar to the general vocabulary of Indonesian, Sketch Engine Corpora, and Leipzig Corpora.

**Keywords:** vocabulary, *Sahabatku Indonesia*, corpus linguistics, BIPA book

### Introduction

In 2024, the global economy is expected to shift its focus from the United States to the Asian region, including countries like China, India, Korea, Japan, and Indonesia. This shift is driven by the significant economic growth projected for these Asian nations, including Indonesia. Indonesia possesses valuable resources, both human and natural, that have the potential to impact the global market. Notably, Indonesia has vast nickel reserves, with an estimated 17.7 billion tons of ore and 177.8 million tons of metal <sup>1</sup>. Nickel is a crucial material for the electronics industry, particularly in nanotechnology. It is used in the production of electric cars, motorbikes, electricity, smartphones, laptops, and various other electronic devices.

Given these circumstances, Indonesia will have the potential to influence the global stage. It is anticipated that numerous foreign countries will rely on Indonesia across various sectors, prompting increased interactions between foreigners and Indonesia. Consequently, more and more foreigners are likely to learn Indonesian through the BIPA program. BIPA, or Indonesian for Foreign Speakers, is a program designed for foreign students who wish to learn Indonesian. Interest in learning Indonesian via BIPA has been notably high; from the start of 2021 until September, there were 8,950 BIPA students from 30 different countries<sup>2</sup>. This growing interest necessitates the development of an

---

<sup>1</sup> Aditya Prabowo, "Kementerian ESDM: Cadangan Nikel Indonesia Masih Terbesar Dunia," *rri.co.id*, 19 Oktober 2023, <https://www.rri.co.id/bisnis/407372/kementerian-esdm-cadangan-nikel-indonesia-masih-terbesar-dunia>.

<sup>2</sup> "Capaian Keberhasilan BIPA Tingkatkan Fungsi Bahasa Indonesia di Kancah Dunia," *Kementerian Pendidikan dan Kebudayaan*, 26 September 2021,

appropriate curriculum and effective teaching methods. A well-crafted curriculum and suitable teaching methods will enhance the effectiveness of learning Indonesian through BIPA.

Badan Bahasa, Kemendikbudristek has anticipated the growing interest of foreigners in learning Indonesian. One initiative to address this is the establishment of a BIPA learning center called BIPA Online, accessible at <https://bipa.kemdikbud.go.id/portal>. The BIPA Online platform offers various learning features, including Learning BIPA, Keeping BIPA, Bakti BIPA, Tebar BIPA, and Tera BIPA. Additionally, in 2019, Badan Bahasa, Kemendikbudristek, published a series of BIPA books "Sahabatku Indonesia" covering levels BIPA 1–7. These books are widely used by BIPA teachers and students for learning Indonesian.

Language plays a central role in learning activities, particularly in language acquisition. Beyond serving as a medium of communication, language itself constitutes the substance and material of learning. One crucial aspect of language learning is the mastery of vocabulary, which forms the foundation of language proficiency. This viewpoint aligns with Wilkins<sup>3</sup> perspective that vocabulary serves as the bedrock of language. For language learners, the initial priority is acquiring a solid vocabulary base. Through understanding vocabulary meanings, students can communicate effectively, albeit with some limitations. Proficiency in vocabulary underpins proficiency in all four language skills: listening, speaking, reading, and writing. According to Tarigan<sup>4</sup>, one's language proficiency expands in tandem with their vocabulary mastery.

Leveled books feature varying vocabularies across different levels, tailored to the curriculum's specifications regarding vocabulary quantity, phonotactic patterns, and language usage. Vocabulary progression in leveled books follows a staged approach, adjusting for the number of words, complexity of phonotactic patterns, word forms, and the minimal mastery required according to each level's learning outcomes outlined in the curriculum. According to Chabibah et al.<sup>5</sup>, adhering to these level-specific criteria ideally enhances students' language skills corresponding to the book levels used. This necessitates that leveled books introduce distinct vocabularies from lower to higher levels.

Analyzing vocabulary becomes more manageable through corpus linguistics. The term "corpus," derived from Latin meaning "body," linguistically refers to "a sample of natural and authentic language data," as defined by McEnery et al.<sup>6</sup>. Consequently, any linguistic research utilizing natural and authentic language samples from the speech community qualifies as corpus linguistics. Within linguistic research development, corpus linguistics functions not as a branch of linguistics like sociolinguistics or psycholinguistics but as a methodological approach, facilitating the processing and analysis of linguistic data extracted from texts for readability and analysis, as noted by Lindquist<sup>7</sup>.

McEnery and Hardie<sup>8</sup> further emphasize that corpus linguistics employs computer-processed manuscripts to make linguistic data accessible for research purposes. They

---

<https://www.kemdikbud.go.id/main/blog/2021/09/capaian-keberhasilan-bipa-tingkatkan-fungsi-bahasa-indonesia-di-kancah-dunia>.

<sup>3</sup> Thornburry Scott, *How to Teach Vocabulary* (New York: Pearsons Education Limited, 2022).

<sup>4</sup> Tarigan Hanry Guntur, *Pengajaran Kosakata* (Bandung: Angkasa, 1993).

<sup>5</sup> Shelya Chabibah dan Kisyani, "Perkembangan Kosakata dan Fonotaktik pada Buku Ajar BIPA Tingkat A1 – C2 Terbitan Kemendikbud," *Jurnal Bapala*, 09, 01 (2018).

<sup>6</sup> Tony McEnery dan Andrew Hardie, *Corpus Linguistics* (UK: Cambridge University Press, 2012).

<sup>7</sup> Hands Lindquist, *Corpus Linguistics and the Description of English* (Edinburgh: Edinburgh University Press, 2009).

<sup>8</sup> McEnery dan Hardie, *Corpus Linguistics*.

assert the impracticality of manually handling corpus data due to its volume and the potential for errors, advocating for computer-based processing to effectively and efficiently manage extensive textual data for descriptive linguistic research.

Research conducted by Jang Ho Lee et al.<sup>9</sup> explored the relationship between vocabulary learning and learning strategies among Korean language learners as a second language (L2). The findings indicate that motivation directly influences the strategies employed for vocabulary acquisition and the level of vocabulary mastery (B2) among students. Higher motivation correlates with increased vocabulary proficiency, facilitating faster development of speaking skills in the second language being studied<sup>10</sup>.

Another study by Kharis<sup>11</sup> investigated entries in leveled books in his dissertation titled "Lemas in Netzwerk Leveled Books Level A1 – B1 Standard Common European Framework of Reference for Languages (CEFR): A Computational Linguistic Study." This research revealed inconsistencies in the density ratio across different levels of Netzwerk books from A1 to B1. The analysis indicated that not all students comprehended the meanings of words used in these books uniformly. While sharing a similar research methodology, this study differs in terms of its research subjects, data, theories, focus, and tools employed, suggesting potentially different research outcomes.

Developing the BIPA curriculum and learning materials requires adherence to several essential principles. These include ensuring relevance by aligning curriculum content with the current and future needs of students through needs assessment analysis, enhancing effectiveness by designing curricula that support learning and goal achievement, promoting efficiency by ensuring learning materials are practical and suitable, ensuring continuity through well-organized program levels and course distributions, and fostering flexibility to empower teachers in enhancing their competencies.

In order to effectively write various types of written work, language learners must acquire a specific range of vocabulary relevant to that field. According to Nation<sup>12</sup>, educators teaching a language should identify the essential types and minimum quantities of vocabulary that students need to master. This approach aims to enhance the efficiency of vocabulary acquisition among language learners, thereby expediting their proficiency in achieving specific language goals. These identified vocabularies serve as the foundation for developing textbooks, learning modules, BIPA dictionaries, and other instructional materials in Indonesian for Foreign Speakers (BIPA).

Research conducted by Dwi Kurniasih<sup>13</sup> examined the BIPA book "Sahabatku Indonesia" at the basic level. The study found that the content and materials in the BIPA book were challenging for comprehension. Consequently, the research suggests the necessity of creating pre-teaching matriculation books aligned with language corpus data. Developing these corpus-based pre-teaching books proves beneficial for refining the

---

<sup>9</sup> Jang Ho Lee, Joung Joo Ahn, dan Hansol Lee, "The Role of Motivation and Vocabulary Learning Strategies in L2 Vocabulary Knowledge: A Structural Equation Modeling Analysis," *Studies in Second Language Learning and Teaching* 12, no. 3 (2022): 435–58.

<sup>10</sup> Lee, Ahn, dan Lee.

<sup>11</sup> M Kharis, "Lema Dalam Buku Berjenjang Netzwerk Level A1 – B1 Standar Common European Framework of Reference for Languages (CEFR): Studi Linguistik Komputasional" (Disertasi, Surabaya, Universitas Negeri Surabaya, 2014).

<sup>12</sup> Nation, *Learning Vocabulary in Another Language* (UK: Cambridge University Press., 2001).

<sup>13</sup> Dwi Kurniasih dan Siti Isnaniah, "Penerapan Bahan Ajar Bahasa Indonesia bagi Penutur Asing (BIPA) 'Sahabatku Indonesia' Tingkat Dasar di IAIN Surakarta," *Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA)* 1, no. 2 (5 Desember 2019): 62, <https://doi.org/10.26499/jbipa.v1i2.1793>.

microstructure of dictionaries, encompassing entries, word classes, definitions, and usage examples<sup>14</sup>.

However, previous studies on vocabulary-focused language learning in BIPA books have overlooked corpus linguistics, resulting in limited linguistic data processing. Given this gap, the present research aims to explore how vocabulary is utilized across levels 1 to 7 in the "Sahabatku Indonesia" book series using corpus linguistics methods. This study seeks to provide a detailed analysis of vocabulary usage specifically within the context of learning Indonesian for foreign speakers. While research on learning Indonesian for foreign speakers has been conducted previously, the application of corpus linguistics to analyze vocabulary in the "Sahabatku Indonesia" book series represents a novel contribution to language learning scholarship, particularly in the field of Indonesian language education for non-native speakers.

## Methods

This study employs mixed research methods, combining both quantitative and qualitative approaches. Quantitative methods typically utilize statistical and mathematical tools for analysis, while qualitative methods rely on logical reasoning and understanding the interpretation of research subjects<sup>15</sup>. In essence, mixed methods research integrates quantitative and qualitative methodologies in a study, aiming to incorporate them at various stages of the research process<sup>16</sup>.

The data source for this research is derived from the "Sahabatku Indonesia" series of BIPA books published by Badan Bahasa, Kemendikbutristek. This series consists of seven levels, ranging from BIPA 1 to BIPA 7, accessible for download on the BIPA LMS page hosted by the Language Agency at <https://bipa.kemdikbud.go.id/portal>.

The research data is initially obtained in electronic PDF format from the "Sahabatku Indonesia" books and subsequently converted to TXT format for processing using the Ant Word Profiler 2.0.0 corpus software<sup>17</sup>, available for free download at <https://www.laurenceanthony.net/software/antwordprofiler/>. The interface of the Ant Word Profiler corpus software<sup>18</sup> is illustrated in Figure 1.

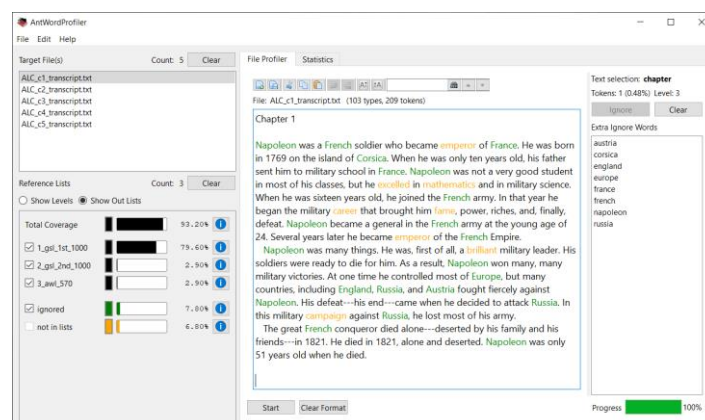


Figure 1: Ant Word Profiler 2.0.0 program display  
(<https://www.laurenceanthony.net/software/antwordprofiler/>)

<sup>14</sup> Kurniasih dan Isnaniah.

<sup>15</sup> Yusrie Abadi, *Penelitian Kualitatif* (Jakarta: Media Group, 2003).

<sup>16</sup> Zoltan Dornyei, *Research Methods in Applied Linguistics* (UK: Oxford University Press, 2007).

<sup>17</sup> Laurence Anthony, "AntWordProfiler" (Japan: Waseda University, t.t.), <https://www.laurenceanthony.net/software/antwordprofiler/>.

<sup>18</sup> Anthony.

The research data is in text format in PDF format and then converted into TXT format so that it can be processed using the Ant Word Profiler 2.0.0 corpus program<sup>19</sup>. Then, the research data is cleaned from lingual elements that are not included in the required research data, such as author identity, publication identity, page numbers, etc.

## Results and Discussion

### Vocabulary in the BIPA Book "Sahabatku Indonesia"

As BIPA students spend more time studying, their language skills and vocabulary mastery improve. This principle applies similarly to leveled language learning books. At the initial stage, students are introduced to a basic set of general vocabulary. As they master these foundational vocabularies, subsequent levels introduce more extensive and specialized vocabularies.

No	Books	Number of Vocabulary	Token
1	BIPA 1	1.966	12.292
2	BIPA 2	1.477	8.754
3	BIPA 3	1.984	12.081
4	BIPA 4	2.440	14.540
5	BIPA 5	3.496	21.948
6	BIPA 6	2.656	16.051
7	BIPA 7	2.973	15.868
8	<b>TOTAL</b>	<b>16.992</b>	<b>101.534</b>

Table 1: Number of Vocabulary and Words in "Sahabatku Indonesia" Book

In the BIPA book "Sahabatku Indonesia" series, the progression of vocabulary across levels does not follow an ideal upward trend. Specifically, the number of vocabulary words varies across different levels. For instance, in BIPA level 1, there are 1,966 vocabulary words used. However, in BIPA level 2, this number decreases to 1,477. Subsequently, BIPA level 3 sees an increase to 1,984 words, followed by BIPA level 4 with 2,440 words, and BIPA level 5 with 3,496 words. In contrast, BIPA level 6 shows a decrease to 2,656 words, and BIPA level 7 sees a slight increase to 2,973 words. This fluctuation in the number of vocabulary words across BIPA levels reflects varying vocabulary requirements that students must master at each educational level.

---

<sup>19</sup> Anthony.

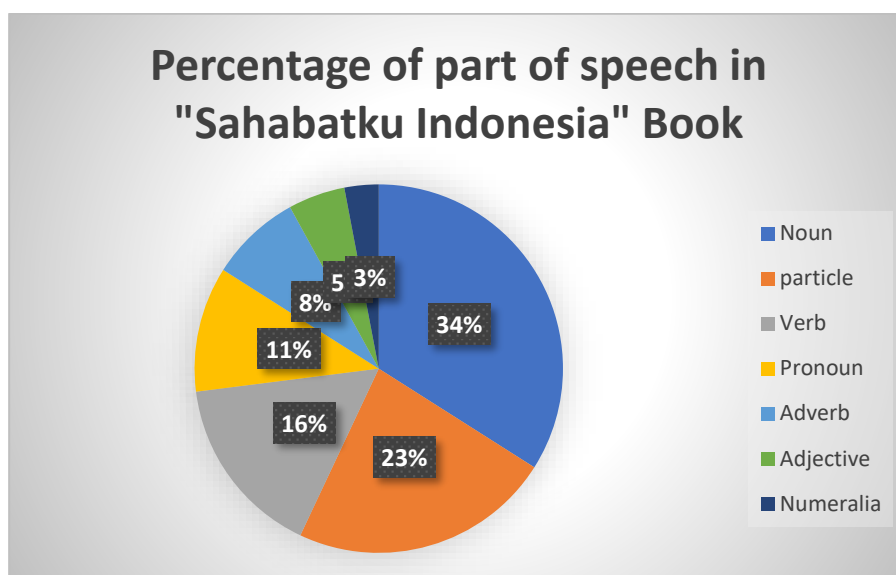


Diagram 1: Percentage of part of speech in "Sahabatku Indonesia" Book

Based on the analysis of the top one hundred vocabularies using Ant Word Profiler 2.0.0 (Anthony, 2022), seven categories of word classes were identified: nouns, particles, verbs, pronouns, adverbs, adjectives, and numerals. The most frequently occurring part of speech is nouns for 34%, followed by particles at 23%, verbs at 16%, pronouns at 11%, adverbs at 8%, adjectives at 5%, and numerals at 3%. The vocabulary words are presented in the table below.

No	Noun	Particle	Verb	Pronoun	Adverb	Adjective	Numeral
1	teks	yang	ada	itu	tidak	baik	banyak
2	kegiatan	dan	mampu	ini	akan	banyak	satu
3	indonesia	di	adalah	anda	juga	lain	beberapa
4	kata	dengan	dapat	saya	sudah	salah	suatu
5	bahasa	dalam	menggunakan	apa	sangat	umum	
6	kalimat	untuk	menjadi	kami	harus	lebih	
7	berikut	pada	bisa	kita	saja		
8	bahan	dari	tersebut	aku			
9	orang	atau	menulis	mereka			
10	audio	tentang	memahami	kamu			
11	hari	ke	merupakan	mana			
12	sumber	oleh	memiliki				
13	contoh	sebagai	kembali				
14	surat	setelah	berdasarkan				
15	unit	karena	berkaitan				
16	hal	secara					
17	struktur	seperti					
18	bipa	kepada					
19	tempat	telah					
20	tahun	ya					
21	rumah	lalu					
22	teman	bagi					
23	informasi	saat					
24	sampah						
25	pertanyaan						
26	ungkapan						
27	masyarakat						
28	ciri						
29	pendapat						
30	materi						

31	kosakata
32	cerita
33	tujuan

Table 2: List of the One Hundred Highest Frequency Part Of Speech in "Sahabatku Indonesia" Book

Looking at the verb column in the table above, it is evident that the verb vocabulary includes both basic verbs and verbs with affixes. Affixed verbs such as those with the prefixes "me-" and "ber-" are notably frequent. Teaching these affixed verbs, particularly "me-" and "ber-," is essential for initial level BIPA students, as verbs often function as predicates in sentences. Verbs account for 16% of the total, indicating that the sentences used in the BIPA "Sahabatku Indonesia" book predominantly feature verb structures.

However, among the top ten most frequent vocabularies in BIPA books 1 to 7, particles or function words emerge as the predominant word class. Particles, which lack lexical meaning and serve primarily grammatical functions<sup>20</sup>, include conjunctions, prepositions, interjections, discourse markers, and affirmative particles in Indonesian. Conjunctions and prepositions are the most frequently appearing particles in BIPA books.

In addition to particles, pronouns also feature prominently among the top ten most frequently appearing words in BIPA books. The specific ten vocabularies that appear most frequently in BIPA books are detailed in the following table:

Vocabulary	Part of Speech	Frequency
yang	conjunction	3206
dan	conjunction	2513
di	preposition	1579
dengan	preposition / conjunction	1325
teks	noun	1077
dalam	preposition/ conjunction	982
itu	pronoun	959
untuk	conjunction	913
ini	pronoun	911
anda	pronoun	803

Table 3: List of the Ten Highest Frequency Vocabulary Words in BIPA Book "Sahabatku Indonesia"

Frequency is the number of occurrences of vocabulary in research data, in this case the BIPA book. Based on the table above, the conjunction "yang" appears 3,206 times and the conjunction "and" appears 2,513 times. The third place is the preposition "di" which appears 1,579 times in the BIPA book. If grouped by word class, the ten vocabularies with the highest frequency consist of conjunctions, prepositions, pronouns and nouns.

High frequency indicates extensive usage of these vocabularies, underscoring their importance for BIPA students' learning. In the BIPA Sahabatku Indonesia book, lesson materials are organized according to language skills: listening, speaking, reading, writing, and grammar. Introduction to conjunctions, prepositions, and pronouns can be integrated into grammar lessons at the foundational level. Proper placement of conjunctions,

<sup>20</sup> Ida Bagus Putrayasa, *Fungsi Kategori dan Analisis Kalimat* (Bandung: PT. Repika Aditama, t.t.).

prepositions, and pronouns is crucial as these words frequently appear throughout the text.

Referring to table 3, the word "kamu" (you) appears 803 times. "Kamu" is a personal pronoun. This vocabulary can be introduced in the first chapter of BIPA 1, which covers self-introduction. The lesson would explore other personal pronouns such as "saya" (I), "kita" (we), "kami" (us), "kalian" (you guys), "dia" (he/she), and others. Analysis also reveals that "saya" (I) is used more frequently than "aku" (I), suggesting that the texts in the book often depict formal situations while occasionally incorporating informal or casual language examples.

### Comparison of the Vocabulary of the BIPA Book "Sahabatku Indonesia" with Sketch Engine Corpora and Leipzig Corpora

No.	Sketch Engine	Leipzig	Buku BIPA
1.	yang	yang	yang
2.	dan	dan	dan
3.	di	di	di
4.	dengan	ini	dengan
5.	ini	itu	teks
6.	untuk	dengan	dalam
7.	dalam	untuk	itu
8.	dari	dari	untuk
9.	itu	tidak	ini
10.	tidak	akan	anda

Table 4: Comparison of Vocabulary Data for "Sahabatku Indonesia" Book, Sketch Engine Corpora, and Leipzig Corpora

Sketch Engine serves as a corpus data repository for various languages, including Indonesian, accessible at [www.sketchengine.eu](http://www.sketchengine.eu). This resource comprises billions of words of original text, aggregated into corpora that can be instantly analyzed through algorithms. Sketch Engine hosts 600 readily available corpora in over 90 languages, each containing approximately 60 billion words, ensuring a comprehensive and representative sample of language usage<sup>21</sup>. Meanwhile, the Leipzig corpus data bank, maintained by Leipzig University in Germany, also includes Indonesian among its collection of 293 languages worldwide. The Indonesian section of the Leipzig corpus consists of 74,329,815 sentences and 1,206,281,985 tokens (word data)<sup>22</sup>.

Examining table 4 above, it is evident that the top ten highest frequency vocabularies across these three corpus data sources are similar. The most frequent words across all three datasets include "yang", "dan," and "di", indicating their frequent use in Indonesian. Each corpus data bank contains distinct language data in terms of text type and sources, yet collectively they offer a realistic portrayal of Indonesian language usage by its speakers. Furthermore, these data banks reveal common vocabularies such as "dengan", "ini", "untuk", "dalam", "dari", "ini", and "itu" albeit at varying frequency levels. This underscores their prevalence in everyday Indonesian discourse.

<sup>21</sup> "Sketch Engine" (Lexical Computing CZ s.r.o., t.t.), [www.sketchengine.eu](http://www.sketchengine.eu).

<sup>22</sup> "Wortschatz Leipzig" (Universität Leipzig, 2023).



## Vocabulary Test Results for the BIPA Book "Sahabatku Indonesia" Using the Indonesian High Frequency Word List

LEVEL	FILE	TOKEN	TOKEN%
1	Kwary_ID_List1_Apr2013.txt	61792	60.86
0	-	39743	39.14.00

Table 5: Vocabulary Test Results for the BIPA Book "Sahabatku Indonesia" Using the Indonesian High Frequency Word List

Based on the test results, the coverage of the Indonesian high frequency word list is 60.86%. This indicates that the vocabulary in the BIPA book "Sahabatku Indonesia" partially aligns with the list of commonly used Indonesian words. The vocabulary employed in "BIPA My Indonesian Friends" consists of Indonesian high frequency word list. Higher scores indicate greater use of general vocabulary, thereby facilitating comprehension for BIPA students. At the foundational level, language learning books should be structured around common vocabulary to achieve high test scores. Introducing learners to commonly used language terms enables them to quickly apply these words in everyday conversations.

### Conclusion

The findings of this research can be summarized as follows:

- [1] The vocabulary used in the BIPA "Sahabatku Indonesia" Book consists of 16,992 vocabularies and 101,534 words (tokens). Each level of the book shows an increase in vocabulary, although there is a decrease in vocabulary in the BIPA 6 and BIPA 7 levels compared to the previous level.
- [2] The most frequently appearing word classes are nouns, particles, verbs, pronouns, adverbs, adjectives, and numeralia (number words).
- [3] Affix verbs such as "me-" and "ber-" are frequently observed in the text.
- [4] Among the ten most frequent vocabularies in BIPA books 1 to 7, particles (or function words) like conjunctions, prepositions, and pronouns are the most prevalent.
- [5] There are similarities in the highest frequency vocabulary data between the BIPA My Indonesian Friends Book and the Sketch Engine and Leipzig Corpora.
- [6] The coverage of the Indonesian high frequency word list is 60.86%, indicating partial alignment with commonly used Indonesian vocabulary.

Based on these findings, it can be concluded that the use of vocabulary in the BIPA book "Sahabatku Indonesia" is generally suitable for BIPA students' learning. However, there is a need to increase both the quantity of vocabulary and the number of words (tokens) at each level from BIPA 1 to BIPA 7 to enhance the effectiveness of the learning process.

Furthermore, corpus linguistic research focusing on Indonesian language subjects should be expanded in terms of quantity and variety. For example, compiling a technical vocabulary list specific to BIPA has not been undertaken in Indonesia. This research would enrich Indonesian language corpus research. Future research endeavors could focus on developing such a technical vocabulary list, which would not only contribute to advancing corpus linguistic studies but also aid in the preparation of books, modules, and BIPA learning materials.

## Acknowledgement

This scientific article was written by Dadang Rhubido, Kisyani Laksono, and Syamsul Sodiq based on the results of research entitled "Vocabulary in the BIPA "Sahabatku Indonesia" Book: A Corpus-Based Study" which was funded by BPPT and LPDP. The content is the sole responsibility of the author.

## Bibliography

- Abadi, Yusrie. *Penelitian Kualitatif*. Jakarta: Media Group, 2003.
- Anthony, Laurence. "AntWordProfiler." Japan: Waseda University, n.d. <https://www.laurenceanthony.net/software/antwordprofiler/>.
- Chabibah, Shelya, and Kisyani Kisyani. "Perkembangan Kosakata dan Fonotaktik pada Buku Ajar BIPA Tingkat A1—C2 Terbitan Kemendikbud." *Jurnal Bapala* 09, no. 01 (2018).
- Dornyei, Zoltan. *Research Methods in Applied Linguistics*. UK: Oxford University Press, 2007.
- Hanry Guntur, Tarigan. *Pengajaran Kosakata*. Bandung: Angkasa, 1993.
- Kementerian Pendidikan dan Kebudayaan. "Capaian Keberhasilan BIPA Tingkatkan Fungsi Bahasa Indonesia di Kancah Dunia." September 26, 2021. <https://www.kemdikbud.go.id/main/blog/2021/09/capaian-keberhasilan-bipa-tingkatkan-fungsi-bahasa-indonesia-di-kancah-dunia>.
- Kharis, M. "Lema Dalam Buku Berjenjang Netzwerk Level A1–B1 Standar Common European Framework of Reference for Languages (CEFR): Studi Linguistik Komputasional." PhD diss., Universitas Negeri Surabaya, 2014.
- Kurniasih, Dwi, and Siti Isnaniah. "Penerapan Bahan Ajar Bahasa Indonesia bagi Penutur Asing (BIPA) 'Sahabatku Indonesia' Tingkat Dasar di IAIN Surakarta." *Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA)* 1, no. 2 (December 5, 2019): 62. <https://doi.org/10.26499/jbipa.v1i2.1793>.
- Lee, Jang Ho, Joung Joo Ahn, and Hansol Lee. "The Role of Motivation and Vocabulary Learning Strategies in L2 Vocabulary Knowledge: A Structural Equation Modeling Analysis." *Studies in Second Language Learning and Teaching* 12, no. 3 (2022): 435–58.
- Lindquist, Hans. *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press, 2009.
- McEnery, Tony, and Andrew Hardie. *Corpus Linguistics*. UK: Cambridge University Press, 2012.
- Nation, I.S.P. *Learning Vocabulary in Another Language*. UK: Cambridge University Press, 2001.
- Prabowo, Aditya. "Kementerian ESDM: Cadangan Nikel Indonesia Masih Terbesar Dunia." *rri.co.id*, October 19, 2023. <https://www.rri.co.id/bisnis/407372/kementerian-esdm-cadangan-nikel-indonesia-masih-terbesar-dunia>.
- Putrayasa, Ida Bagus. *Fungsi Kategori dan Analisis Kalimat*. Bandung: PT. Repika Aditama, n.d.
- Scott, Thornbury. *How to Teach Vocabulary*. New York: Pearson Education Limited, 2022.
- "Sketch Engine." Lexical Computing CZ s.r.o., n.d. [www.sketchengine.eu](http://www.sketchengine.eu).
- "Wortschatz Leipzig." Universität Leipzig, 2023.