LINGUISTIK : Jurnal Bahasa & Sastra
http://jurnal.um-tapsel.ac.id/index.php/Linguistik | Vol.10 No.2 April-Juni 2025
e- ISSN 2548 9402 | | DOI : 10.31604/linguistik.v10ii2.258-268

A Corpus-Based Study of High-Frequency Words in Madurese Textbooks for Elementary School

Fitriyatuz Zakiyah, Eka Susylowati
Email: fitriyatuz.zakiyah@trunojoyo.ac.id
Universitas Trunojoyo, Jl. Raya Telang, Perumahan Telang Inda, Telang, Kec. Kamal, Kabupaten Bangkalan, Jawa Timur

Abstrak

Penelitian ini bertujuan untuk menganalisis kata-kata frekuensi tinggi yang digunakan dalam buku teks bahasa Madura untuk siswa kelas 4, 5, dan 6 sekolah dasar.Data dalam penelitian ini adalah korpus yang dibuat dari dua buku teks bahasa Madura untuk sekolah dasar yang digunakan di kelas 4, 5, dan 6 di Bangkalan.Data tersebut diproses menggunakan perangkat lunak konkordansi, yaitu Lancsbox, untuk mengidentifikasi kosakata yang sering digunakan dalam buku teks tersebut.Kemudian, data tersebut dianalisis secara kualitatif untuk mengetahui makna dan kelas kata dari kata-kata tersebut.Hasil penelitian ini menunjukkan bahwa frekuensi kata-kata yang paling sering digunakan dalam buku teks di ketiga kelas adalah sama, yaitu "dan" dan "sè" yang berarti "yang".Selain itu, dalam buku teks Kelas 4, dari 100 kata frekuensi tinggi di seluruh kelas dalam data, terdapat 7 jenis kata yang paling sering digunakan.Itu adalah kata benda, partikel, kata ganti, kata kerja, kata sifat, angka, dan kata keterangan.Dalam buku teks kelas 5 dan 6, bagian dari ucapan yang sering digunakan adalah kata benda.Ini juga menggunakan partikel, kata ganti, kata kerja, kata sifat, kata keterangan, dan angka.pentingnya desain teks yang mendukung transisi dari pembelajaran naratif sederhana di kelas awal ke pembelajaran yang lebih padat informasi dan berbasis fakta di kelas-kelas selanjutnya.Ini menunjukkan pentingnya desain teks yang mendukung proses pembelajaran.Selain itu, hasil ini dapat berfungsi sebagai panduan bagi pendidik dan pengembang kurikulum dalam menciptakan bahan ajar yang lebih efektif, selaras dengan perkembangan bahasa dan kognisi siswa.

Kata kunci: Buku ajar, Kata berfrekuensi tinggi, Korpus, Madura, SD

Abstract

This research aims to analyze high-frequency words used in Madurese language textbooks for grade 4th, 5th, and 6th of elementary school students. The data in this study is a corpus created from two Madurese language textbooks for elementary schools used in grade 4th, 5th, and 6th in Bangkalan. The data was processed using concordance software, namely Lancsbox, to identify the vocabulary that are frequently used in those textbooks. Then, the data was analysed qualitatively to know the meaning and the word classes of the words. The results of this study show that the frequency of words that are most often used in textbooks in the three classes is the same, those are and “sè” which mean "yang". Additionally, in the Grade 4th textbook, out of 100 high-frequency words across the grades in the data, there are 7 parts of speech that are most frequently used. Those are nouns, particles, pronouns, verbs, adjectives, numerals, and adverbs. In the Grade 5th and 6th textbook, the part of speech that frequently used is nouns. It also used particles, pronouns, verbs, adjectives, adverbs, and numerals. the importance of text design that supports the

transition from simple narrative learning in early grades to more information-dense and fact-based learning in later grades. It implies the importance of text design that supports the learning process. Moreover, these results can serve as a guide for educators and curriculum developers in creating more effective teaching materials, aligned with the development of language and student cognition.

Keywords: Corpus, elementary school, high-frequency words, madurese language, textbook.

INTRODUCTION

Madurese language is one of the regional languages in Indonesia that is used by Madura ethnic residents, both those who live in Madura Island and outside Madura Island in daily communication (Sofyan, 2010). The Madurese language become the third out of the ten most used regional languages in Indonesia and the fifth among the ten most populous ethnic groups in Indonesia (Ifada et al., 2023).

As a local language, the Madurese language has three functions. Those are symbol of regional pride, regional identity, and domain of communication within the family and regional community (Effendi, 2011). Therefore, the preservation of the Madurese Language must be directed to its preservation and development, so that this language can provide valuable contributions to the development of national culture.

However, Indonesian language as a national language actually influences the existence of Madurese Language in Madura island, especially in Bangkalan (Rainy, 2015). the geographical location of Bangkalan especially, Kamal, which is close to Surabaya influences the language used in that area. Most people there used Indonesian language when speaking with immigrants (Zakiyah, 2023). The existence of the Madurese language is increasingly worrying, especially the most polite variationof Madurese language, which is the engghi-bhunten or also called the andhep ashor (a language that is polite) (Rainy, 2015). In addition, the geographical location of Madura which is close to Surabaya especially Bangkalan regency has a heterogeneous community. So that, in daily communication, local people more often use Indonesian or even Javanese than Madurese language, such as areas near campuses and housing and public spaces (Zakiyah et al., 2024).

The extinction of regional languages has become a global issue. According to UNESCO, almost every week a regional language becomes endangered language (Sukamto & Qalyubi, 2022). Therefore, efforts are needed to address this issue. One of the ways to protect the Madurese Language is to revitalize it. Grenoble defines revitalization as a social transformation, which not only involves individuals, but also a community of language users (Olko & Sallabank, 2021). Local language revitalization is a way to preserve and develop the languages through the inheritance of the languages to the younger generation. The revitalization of local languages is important thing to maintain cultural and language diversity in Indonesia (Andina, 2023). Thus, language

revitalization is a very important activity to be done to preserve regional languages for the younger generation.

There are 3 language revitalization models. One of them is education-based revitalization. This revitalization model leads to an increase in the mastery of local languages and literature through the field of education such as making the Madurese language one of the local contents in schools in Madura starting from elementary school.

Education-based revitalization is used for regions with vulnerable languages because they are used in competition with other regional languages. Although the number of language speakers is still relatively large, but the use is decreasing (Andina, 2023). The Balai bahasa jawa Timur said that model B or education-based language revitalization is very suitable for revitalizing Madurese Language (Trisna, 2023)

However, the problem comes when students from outside Madura learn the Madurese language at their schools because it is not their local language or their first language. Second language (L2) learners have less exposure to the target language and less learning time compared to children learning their first language (Dang et al., 2020). Therefore, every elementary schools will use Madurese language to help them.

Textbooks are essential tools in language learning class, especially in local language class, such as Madurese language at the elementary school level. It provides structured input that aligns with curriculum goals and student needs. According to Parys et al (2024), textbooks are important source of language input for students. This is in line with De Wilde et al (2022) who stated that textbooks are considered the primary source of language input.

Moreover, textbooks help elementary school students to understand Madurese language vocabulary correctly and properly, especially in writing. This is because students are accustomed to recognizing the Madurese language as a spoken language, so they do not understand how to write several vocabulary correctly and the use of its vocabulary according to the context, especially for students from outside Madura Island.

In the process of language learning, textbooks can serve as a source of language input. Input in language learning is very important, especially for learning a second or foreign language (P. Nation, 2007). Language input in the form of vocabulary can be received by someone in two ways, namely formal and informal. Formally, input is received by someone by engaging in language learning activities in the classroom. As for the informal way, it involves watching TV and playing games outside of class (Peters et al., 2019).

Textbooks as a primary resource can also develop reading skill. To improve students' reading comprehension in language classes, particularly foreign languages, the role of textbooks used in schools must be considered (Rustan & Andriyanti, 2021). The quality and content of textbooks, especially the vocabulary they present, significantly influence the effectiveness of language learning. To optimize language learning, vocabulary in language textbooks must be carefully selected (Rustan & Andriyanti, 2021). For mastering language, learners need to receive substantial input (Webb & Nation, 2017). However, some language textbooks have quite limited input (Jordan & Gray, 2019). The

vocabulary utilized to be fluent reader should be approximately 3000 words (I. S. P. Nation, 2001). These terms can be used as a basis in giving the new vocabulary in the class. Therefore, vocabulary used in the textbook around 2000-3000 words.

In addition, Identifying the vocabulary that L2 learners should learn first is extremely crucial, because it helps them receive the highest return on their learning effort. Therefore, identifying high frequency words in textbooks used in the class is crucial thing. Dang et all (2020) stated that knowledge of high-frequency words is significant because it may help students recognize a big proportion of terms in various spoken and written sources. Moreover, Students knew more high-frequency words than those with lower frequency levels.

HFW is believed to be important, supporting the success of second language learning because learning important words increases the motivation and confidence of learners in producing their own sentences (Siagian, 2020). The higher the learning motivation, the better the results obtained (Maharani, 2019). Moreover, the use of HFW in language teaching is very important, as it is a direct teaching strategy and a method of repetition (Johns & Wilke, 2018). Therefore, analysis of high frequency words in madurese textbook is important.

Several researchers have analyzed high-frequency words in language learning textbooks. Rustan & Andriyanti (2021) analyzed High Frequency Words in English Textbooks for Indonesian Senior High Schools. The analysis results showed that there were 124 words found as the HFWs. The HFWs found were mostly articles, prepositions, pronouns, nouns, verbs, adverbs, adjectives, and conjunctions (Rustan & Andriyanti, 2021). Additionally, Widodo et al. (2022) analyzed high-frequency words in basic level BIPA textbooks. The study showed that the type of low frequency words had high percentage in basic level BIPA textbooks (Widodo et al., 2022) However, from previous studies, research on high-frequency words in local language textbooks, particularly the Madurese language, is very limited. Therefore, in this study, we focuses on analyzing high-frequency words in Madurese language textbooks for elementary school levels. The aims of this research is to identify high-frequency words in Madurese language textbooks for grades 4th, 5th, and 6th of elementary school. Additionally, this study also aims to determine the parts of speech of high-frequency words in Madurese language textbooks for grades 4th, 5th, and 6th of elementary school.

METHOD

This study used descriptive qualitative methods in analyzing the data. It described the high frequency words used in Madurese textbook for elementary school. We use that method because it emphasizes the search for meaning, understanding, concepts, characteristics, symptoms, symbols, and descriptions of a phenomenon using various approaches, and is presented narratively.

The data in this research is corpus data constructed from readings in several Madurese language textbooks for grades 4th, 5th, and 6th under the title Sekkar Assrè and Songsong Sènom. Moreover, in this research, we described the word classes of the high frequency words used in the data. Therefore, we employed descriptive qualitative research. In analyzing the data of qualitative research, researchers would not use statistic methods or formula (Dornyei, 2007).

When collecting the data, we employed concordance software, Lancsbox, to investigate the amount of words. Then we used several steps to collect the data. First, we gathered all of the reading materials from the textbooks. Second, we searched the data for high-frequency words using Lancsbox's wordlist menu (see figure 1).

Fig. 1. Wordlist of the data in Lancsbox

According to data, the readings from the fourth-grade elementary school textbook contain a total of 5,298 words. Meanwhile, the fifth-grade textbook contains a total of 4,623 words. In the sixth grade, a total of 4,124 words are used. The last, we separated out the high frequency words. We sorted the high frequency words from all of the wordlists we identified. There were just 300 words in the data.

The data were then processed to determine the word classes of the most frequently used words in the dataset. First, we divided the words we discovered depending on their frequency. Second, we classified the terms we found based on their word classifications. Finally, we examined concordance to identify the meaning and its part of speech (figure 2).

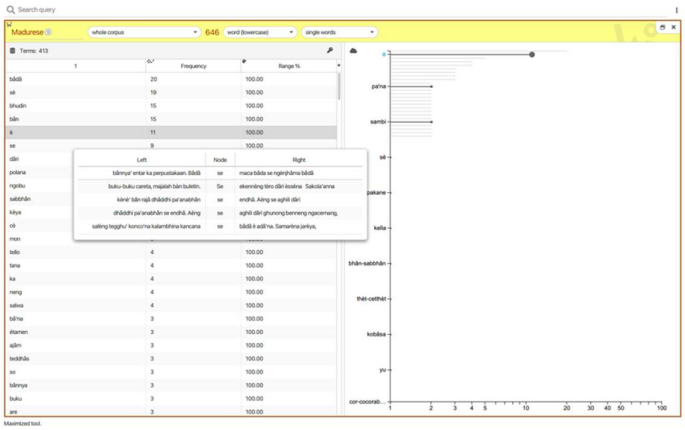

Fig. 2. Concordance in Lancsbox

RESULTS

Data showed that the token of 4th grade was 13,723, 5th grade was 10,718, while 6th grade was 10,851. Moreover, there were a total of 5,298 types of words in the readings in the 4th grade elementary school textbook. While in the 5th grade book, there were a total of 4,623 types of words.

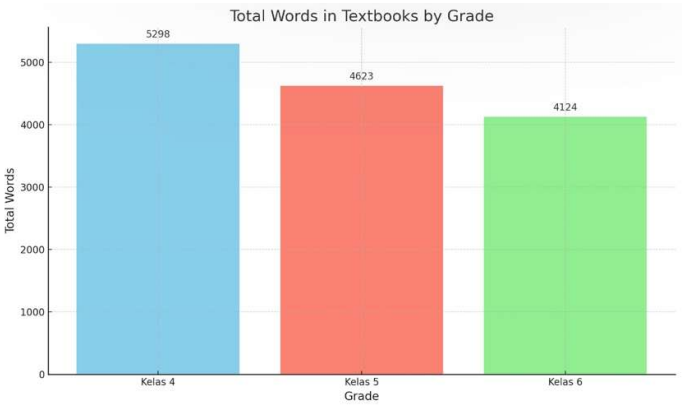

Fig. 3. Total types of words in the data

As for 6th grade, the total types words used were 4,124 (see figure 3). The highest frequently word used in the textbooks in the three classes was same. It was “sè” which mean "yang" (see table 1).

Table 1
Five high frequency words used in the data

4th Grade		5th Grade		6th Grade	
Words	Frequency	Words	Frequency	Words	Frequency

sè	624	sé	sé	Sè	192
panèka	387	paneka	217	Paneka	189
ka	239	sareng	178	Sareng	185
sareng	192	ka	98	engghi	64
orèng	126	kalabân	73	neng	61

In the 4th Grade textbook, out of 100 high-frequency words across the grades in the data, there are 6 part of speech that are most frequently used. Those were particles 20 words, nouns 38 words, pronouns 7 words, verbs 16 words, adjectives 6 words, and adverbs 13 words. In the 5th Grade textbook, there were particles 20 words, nouns 39 words, pronouns 6 words, verbs 16 words, adjectives 7 words, and adverbs 8 words and numerals 4 words. Meanwhile, in the 6th Grade textbook, particles was 14 words, nouns was 46 words, pronouns was 2 words, verbs was 15 words, adjectives was 8 words, adverbs was 13 words, and numerals 2 words (see figure 4).

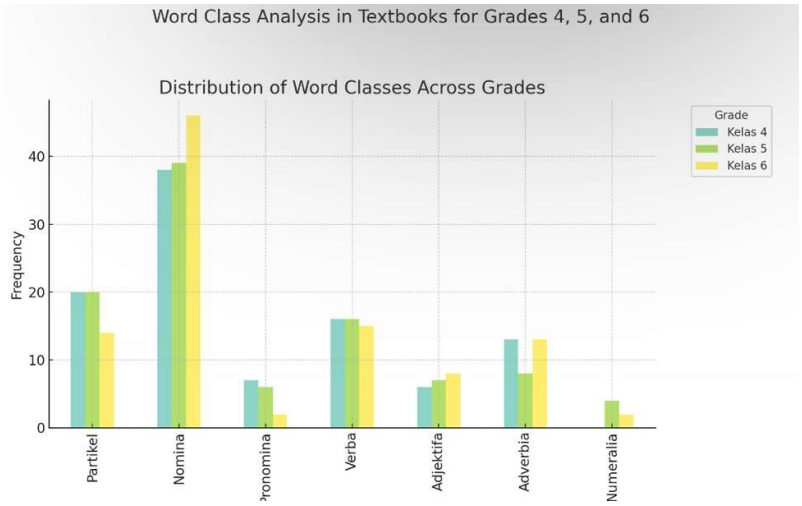

Fig. 4. The distribution of word classes across the grades

DISCUSSION

The high frequency words in Madurese Textbooks

The research results showed that the tokens of 4th grade was 13,723, 5th grade was 10,718, while 6th grade was 10,851. The total types of words used in the Madurese language textbook for 4th grade elementary school is 5,298 words. This indicates that at this level, the texts in the textbooks tend to be longer and more descriptive. The focus of learning may be more on exploring stories or simple narratives that require more words.

This is different from the textbooks in 5th grade. The total types of word count decreases to 4,623 words. This decrease can be linked to the change in text types to more concise ones, such as information-oriented texts or denser descriptions. Meanwhile, the word count in the 6th grade textbooks is even lower, at 4,124 words.

This decrease reflects a shift in the function of the presented texts, which are now more focused on the efficient delivery of ideas or facts. The decrease in the total types word count from 4th grade to grade 6th indicates a change in the strategy of presenting

learning materials. Textbooks for higher grades tend to use more goal-oriented texts, with more complex sentence structures but fewer words. In addition, 4th grade students are still in the stage of understanding longer readings with simple vocabulary. In 5th and 6th grades, the reading material is denser with information, making the texts shorter but more meaningful.

At all three levels, the most frequently used word is “sè”, which means "yang". It is in line with the Indonesian language, which the word "yang" becomes the most frequently used word in Indonesian textbook of seventh, eighth, and ninth grades was “yang” (Solihah et al., 2021). It is a relative conjunction and is often used to explain or provide further information in a sentence. The word " sè ", in the context of textbooks, this word may often appear in descriptive sentences, such as

*“Puisi **sè** aghandhu' carêta orèng lakè' tarèsna dâ' orêng bine' otabâ sabhâlikkèpon”*

means

*“a poem **that** tells the story of a man who loves/likes a woman or vice versa”*

The presence of this word " sè " in the sentence above is very important because it helps to construct logical and informative sentences, especially in texts that function to describe or explain something.

Moreover, the frequent use of the word "sè" at all levels indicates that although the total number of words has decreased, the role of functional words remains dominant. This reflects the nature of the language of instruction in education, which requires sentence structures that are easy for students to understand while still providing logical relationships between the elements of the sentence.

Distribution of word classes in the data

Based on the research findings, from 100 words in the textbooks of each class, 7 parts of speech are most frequently used, namely particles, nouns, pronouns, verbs, adjectives, adverbs, and numerals. It is in line with BIPA textbook for basic level. The word classes in those textbooks was vary (Siagian, 2020). In Grades 4th and 5th, the frequency of particles was 20 words. This shows that the use of functional words such as "ka" and "sè," is still an important part of text composition for students at this level. This stability may be related to the basic function of particles in constructing simple sentences.

It is different from grade 6th, where the frequency of preposition words decreases to 14 words. This decrease may be due to the replacement of preposition functions with phrases or more complex sentence structures, in line with the development of students' language abilities, which are beginning to focus on variations in grammatical usage.

The second most frequently used part of speech is nouns. There are 38 words classified as nouns in the 4th -grade textbook. This shows that focus on introducing vocabulary in the form of nouns such as names of objects, places, or concrete concepts, which are important in early learning, such as the word “orèng” means “people”, and the word “madhurâ” means Madura or indicates place. In the 5th grade textbook, the number of nouns slightly increased to 39 words. This increase indicates an expansion of the students' vocabulary with the addition of abstract nouns or more complex nouns, such as the word "pangator," which means organizer or determiner, and "saloka," which means action. As for the 6th grade textbooks, a significant increase occurred in the use of nouns. Those are 46 words. It indicates that at this level, students are introduced to texts that are

denser with nouns, such as historical narratives.

CONCLUSION

From this study, it can be concluded that Madurese language textbooks for grades 4th to 6th experience a gradual decrease in the total number of word types. The grade 4th textbook contains longer and more descriptive texts to support students' understanding of simple stories. In contrast, the grade 5th and 6th textbooks tend to present more concise and information-dense texts. This change reflects a shift in the strategy of material presentation from simple narrative-based learning to more goal-oriented material with more complex sentence structures. Moreover, the word "sè," which means "that," has become the most frequently used word at all grade levels. This word reflects the importance of the role of functional words in building logical relationships between the elements of a sentence. Its use aligns with the nature of instructional language, which must be easily understood by students while still providing a logical structure. Furthermore, the distribution of word classes also shows significant changes. Particles dominate in grades 4th and 5th, but their frequency decreases in grade 6th as the complexity of sentence structures increases. Meanwhile, the number of nouns increases significantly from grade 4th to grade 6th, reflecting the introduction of more complex concepts. This indicates that the Madura language textbooks are designed to support students' language development, from understanding basic concepts to the ability to process denser information with a wider grammatical variation. This implies the importance of text design that supports the transition from simple narrative learning in early grades to more information-dense and fact-based learning in later grades. These results can serve as a guide for educators and curriculum developers in creating more effective teaching materials, aligned with the development of language and student cognition.

RECOMMENDATION

The recommendation for further research can conduct a comparative analysis of Madurese language textbooks with other regional language textbooks to observe the distribution patterns of words and different teaching approaches. Additionally, future research can also explore the use of technology, such as interactive learning applications, to support the teaching of Madurese, particularly in vocabulary mastery and sentence structure. With this recommendation, future research is expected to strengthen the efforts to preserve the Madurese language while also improving the quality of learning in elementary schools.

REFERENCES

Andina, E. (2023). Implementasi dan Tantangan Revitalisasi Bahasa Daerah di Provinsi Lampung. *Aspirasi: Jurnal Masalah-Masalah Sosial*, *14*(1). https://doi.org/10.46807/aspirasi.v14i1.3859

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, *26*(4), 617–641. https://doi.org/10.1177/1362168820911189

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2022). FORMAL VERSUS INFORMAL L2 LEARNING: HOW DO INDIVIDUAL DIFFERENCES AND WORD-RELATED VARIABLES INFLUENCE FRENCH AND ENGLISH L2 VOCABULARY LEARNING IN DUTCH-SPEAKING CHILDREN? *Studies in Second Language Acquisition*, *44*(1), 87–111. https://doi.org/DOI: 10.1017/S0272263121000097

Dornyei, Z. (2007). *Research Methods in Applied Linguistics. Oxford: Oxford University Press.*

Effendi, M. H. (2011). Tinjauan Deskriptif Tentang Varian Bahasa Dialek Pamekasan. *Okara*, *1*.

Ifada, N., Rachman, F. H., Syauqy, M. W. M. A., Wahyuni, S., & Pawitra, A. (2023). MadureseSet: Madurese-Indonesian Dataset. *Data in Brief*, *48*, 109035. https://doi.org/10.1016/j.dib.2023.109035

Johns, J. L., & Wilke, K. H. (2018). HIGH-FREQUENCY WORDS: SOME WAYS TO TEACH AND HELP STUDENTS PRACTICE AND LEARN THEM. In *Texas Journal of Literacy Education* | (Vol. 6, Issue 1).

Jordan, G., & Gray, H. (2019). We need to talk about coursebooks. *ELT Journal*, *73*(4), 438–446. https://doi.org/10.1093/elt/ccz038

Maharani, A. V. (2019). Pemerolehan Kosakata Bahasa Korea pada Pembelajar Dewasa Indonesia. *Ranah: Jurnal Kajian Bahasa*, *8*(2), 255. https://doi.org/10.26499/rnh.v8i2.962

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.

Nation, P. (2007). The Four Strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13. https://doi.org/10.2167/illt039.0

Olko, J., & Sallabank, J. (2021). *Revitalizing Endangered Languages: A Practical Guide*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781108641142

Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The Impact of Instruction and Out-of-School Exposure to Foreign Language Input on Learners' Vocabulary Knowledge in Two Languages. *Language Learning*, *69*(3), 747–782. https://doi.org/https://doi.org/10.1111/lang.12351

Rainy, A. (2015). Pergeseran Penggunaan Bahasa Madura di Kalangan Anak- anak Sekolah Dasar Negeri di Desa Pangarangan Kecamatan Kota Sumenep. *Okara*.

Rustan, R. M., & Andriyanti, E. (2021). High frequency words in english textbooks for indonesian senior high schools. *Studies in English Language and Education*, *8*(1), 181–196. https://doi.org/10.24815/siele.v8i1.18141

Siagian, E. N. (2020a). Kata Berfrekuensi Tinggi dalam Pembelajaran BIPA Pemula. *Ranah: Jurnal Kajian Bahasa*, *9*(2), 188. https://doi.org/10.26499/rnh.v9i2.2320

Siagian, E. N. (2020b). Kata Berfrekuensi Tinggi dalam Pembelajaran BIPA Pemula. *Ranah: Jurnal Kajian Bahasa*, *9*(2), 188. https://doi.org/10.26499/rnh.v9i2.2320

Sofyan, A. (2010). Fonologi Bahasa Madura. *Humaniora*, *Volume 22*, 207–218.

Solihah, A., Rasyid, Y., Attas, S. G., & Firdaus, W. (2021). Vocabulary Knowledge Students of Indonesian Language Text Books. *Ilkogretim Online - Elementary Education Online, 20*(1). https://doi.org/10.17051/ilkonline.2021.01.55

Sukamto, K. E., & Qalyubi, I. (2022). *Pedoman Revitalisasi Bahasa Daerah Model B*. Badan Pengembangan dan Pembinaan Bahasa.

Trisna. (2023, May 23). Tergolong Rentan! Revitalisasi Bahasa Madura Pakai Model B: Berbasis Sekolah dan Komunitas. *Portaltiga*.

Van Parys, A., De Wilde, V., Macken, L., & Montero Perez, M. (2024). Vocabulary of reading materials in English and French L2 textbooks: A cross-lingual corpus study. *System*, *124*, 103396. https://doi.org/https://doi.org/10.1016/j.system.2024.103396

Webb, S., & Nation, P. (2017). *How Vocabulary is Learned*. Oxford University Press. https://books.google.co.id/books?id=OwkjvgAACAAJ

Widodo, M., Destiani, & Rudy, M. (2022). Scrutinizing Vocabulary Input In The Basic Class Of Indonesian Language For Foreign Learners: A Corpus Study. *Journal of Positive School Psychology*.

Zakiyah, F. (2023). *Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA) High-frequency affixed words in BIPA 3 textbooks: a corpus-based study*. *5*, 23–32. https://doi.org/10.26499/jbipa.v5i1.6157

Zakiyah, F., Susylowati, E., & Shabrina, K. N. (2024). Language used in shop signs in Kamal, Madura: Virtual landscape linguistics using google street view. In A. Ma'arif (Ed.), *E3S Web of Conferences* (p. 01022). https://doi.org/10.1051/e3sconf/202449901022