A CORPUS STUDY OF ENGLISH VOCABULARY IN TEACHERS CONSTRUCTED TEXTBOOK BY ENGLISH TEACHER FORUM OF JUNIOR HIGH SCHOOL IN TUBAN

THESIS

Submitted in partial fulfillment of the requirement for the degree of Sarjana

Pendidikan (S.Pd) in Teaching English



By:

Siti Uswatun Khasanah

NIM: D05217021

ENGLISH LANGUAGE EDUCATION DEPARTMENT
FACULTY OF TARBIYAH AND TEACHER TRAINING
UIN SUNAN AMPEL SURABAYA

PERTANYAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Siti Uswatun Khasanah

NIM : D05217021

Program studi: Pendidikan Bahasa Inggris

Fakultas : Tarbiyah dan Keguruan

Dengan ini menyatakan bahwa tugas akhir (skripsi) yang berjudul: A Corpus Study of English Vocabulary in Teachers Constructed Textbook by English Teacher Forum of Junior High School in Tuban, adalah benar-benar hasil karya saya sendiri bukan hasil pengambilan tulisan atau karya orang lain yang saya akui sebagai tulisan saya. Semua kutipan yang diperoleh dari sumber lainnya telah disertai keterangan mengenai identitas sumbernya.

Demikian yang dapat saya sampaikan, apabila penyataan tidak sesuai dengan fakta yang ada, maka saya selaku penulis bersedia dimintai pertanggung jawaban.

Surabaya, 3 Juli 2021

Pembuat pernyataan

Siti Uswatun Khasanah

D05217021

422B7AJX595300238

ADVISOR APPROVAL SHEET

This thesis by Siti Uswatun Khasanah entitled "A Corpus Study of English Vocabulary in Teachers Constructed Textbook by English Teacher Forum of Junior High School in Tuban" has been approved by thesis advisors for further approval by the board examiners.

Surabaya, 24 June 2021 Advisor I,

Dr. Siti Asmiya<mark>lı, M.TESO</mark>l

NIP: 197704142006042003

Advisor II,

Rizka Salriyani, M.Pd.

NIP: 198409142009122005

EXAMINER APPROVAL SHEET

This thesis by Siti Uswatun Khasanah entitled "A Corpus Study of English Vocabulary in Teachers Constructed Textbook by English Teacher Forum of Junior High School in Tuban" has been examined on 8 July 2021 and approved by the

board examiners.

Dean,

NIP. 196301231993031002

Examiner I,

H. Mokhamad Syaifudin, M.Ed., Ph.D

NIP. 197310131997031002

Examiner II,

<u>Dr. Hyna Soraya, M.P.d</u> NIP: 196709301993032004

Examiner III,

Dr. Siti Asmiyat, M. TESOI

NIP: 197704142006042003

Examiner IV,

Rizka Safriyani, M.Pd.

NIP: 198409142009122005



KEMENTERIAN AGAMA UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300 E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akad	emika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:
Nama	Siti Uswatun Khasanah
NIM :	D05217021
Fakultas/Jurusan	Pakultas Tarbiyah dan Keguruan / Pendidikan Bahasa Inggris
E-mail address	sitiuswatunkha@gmail.com
UIN Sunan Ampel S	an ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah: 'Tesis 🗆 Desertasi 🗀 Lain-lain()
	OY OF ENGLISH VOCABULARY IN TEACHERS CONSTRUCTED INGLISH TEACHER FORUM OF JUNIOR HIGH SCHOOL IN TUBAN
Perpustakaan UIN mengelolanya dal menampilkan/mem akademis tanpa pe penulis/pencipta da	yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, am bentuk pangkalan data (database), mendialih-mediakannya, dan publikasikannya di Internet atau media lain secara fulltext untuk kepentingan du meminta ijin dari saya selama tetap mencantumkan nama saya sebagai n atau penerbit yang bersangkutan. k menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN
	aya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta
Demikian pemyataa	n ini yang saya buat dengan sebenamya.
	Surabaya, 8 Juli 2021
	Penulis
	U Sunt

(Siti Uswatun Khasanah)

ABSTRACT

Khasanah, S.U. (2021). A Corpus Study of English Vocabulary in Teachers Constructed Textbook by English Teacher Forum of Junior High School in Tuban. A Thesis. English Language Education Department, Faculty of Tarbiyah and Teacher Training, UIN Sunan Ampel Surabaya. Advisor: Dr. Siti Asmiyah, M. TESOL, and Rizka Safriyani, M.Pd.

Keywords: corpus, vocabulary, English textbook

Over the last decade, there has been an increase in corpus-based studies and English Language Teaching (ELT) research. Moreover, there are some researches about corpus study, but it has not specifically discussed the textbooks. This study employs corpus-based analysis to focus on the frequency of English vocabulary used by the English Teachers forum to construct English textbooks in Indonesia, especially in the EFL context. Five textbooks for Junior High School in Tuban have been analyzed using AntWordProfiler 1.4.0 free software from Anthony Laurence. The corpus analysis of selected core textbooks contains 82,918 tokens and 6,669 types. Then, the most frequent words used were those included in the category of High Frequently Word, namely nouns, verbs, and determiners. It should be prioritized in the target words to facilitate students' vocabulary development in the Indonesian EFL context, particularly those in the Tuban district. Besides, the textbooks contained 282-word families of academic words and accounted for about 2.51% of the tokens. In the end, the current research provides teachers with an understanding of which vocabulary they need to teach students to meet the curriculum expectations, and then the English language teaching will be more effective.

ABSTRAK

Khasanah, S.U. (2021). A Corpus Study of English Vocabulary in Teachers Constructed Textbook by English Teacher Forum of Junior High School in Tuban. Skripsi. Pendidikan Bahasa Inggris, Fakultas Tarbiyah dan Keguruan, UIN Sunan Ampel Surabaya. Pembimbing: Dr. Siti Asmiyah, M. TESOL, dan Rizka Safriyani, M.Pd.

Kata kunci : korpus, kosakata, buku teks bahasa Inggris

Selama dekade terakhir, telah terjadi peningkatan studi berbasis korpus dan penelitian tentang pengajaran bahasa Inggris. Selain itu, terdapat penelitian tentang studi korpus tetapi tidak secara khusus membahas tentang buku teks. Penelitian ini menggunakan analisis berbasis korpus yang berfokus pada frekuensi kosakata bahasa Inggris yang digunakan oleh forum guru bahasa Inggris untuk membangun buku teks bahasa Inggris di Indonesia, terutama dalam konteks bahasa Inggris sebagai bahasa asing. Lima buku teks untuk sekolah menengah pertama di Tuban telah dianalisis menggunakan perangkat lunak AntwordProfiler 1.4.0 dari Anthony Laurence. Analisis korpus dari buku teks berisi 82.918 token dan 6.669 jenis kata. Kemudian, kata-kata yang paling sering digunakan yang termasuk dalam kategori kata yang sering muncul, yaitu kata benda, kata kerja, dan determiner. Hasil penelitian ini sangat diprioritaskan dalam kata-kata target untuk memfasilitasi perkembangan kosa kata siswa dalam konteks bahasa Inggris sebagai bahasa asing di Indonesia, terutama yang ada di Tuban. Selain itu, buku teks berisi 282 kata katakata akademik dan menyumbang sekitar 2,51% dari keseluruhan token. Pada akhirnya, penelitian saat ini memberikan guru pemahaman tentang kosa kata yang mereka butuhkan untuk mengajar siswa dalam memenuhi kebutuhan kurikulum, serta menjadikan pengajaran bahasa Inggris akan lebih efektif.

TABLE OF CONTENT

PERNYATAAN KEASLIAN TULISAN	i
ADVISOR APPROVAL SHEET	
EXAMINER APPROVAL SHEET	
LEMBAR PERNYATAAN PUBLIKASI PERPUSTAKAAN	
ABSTRACT	
TABLE OF CONTENTS	
LIST OF TABLES	
LIST OF FIGURES	
LIST OF APPENDICES	
LIST OF ABBREVIATIONS	
CHAPTER I : INTRODUCTION	
A. Background of the Study	
B. Research Question	
C. Objectives of the Study	
D. Significance of the Study	
E. Scope and Limits of The Study	
F. Definition of Key Terms	
CHAPTER II : REVIEW OF RELATED LITERATURE	
A. Theoritical Framework	
1. Corpus and Corpus Analysis	
2. Vocabulary	
3. The use of Corpora in English Language Teaching (ELT)	
4. Textbook	
B. Previous Studies	24
CHAPTER III : RESEARCH METHOD	28
A. Research Design	
B. Subject and Setting of the Study	
C. Data and Source of Data	
D. Research Instrument	30
E. Data Collection Technique	
F. Data Analysis Technique	
G. Checking Validity of Finding	36
CHAPTER IV : RESEARCH FINDING & DISCUSSION	
A. Research Finding	3/
1. Types and Tokens of the Textbooks	
2. Frequency of Vocabulary based on the Part of Speech	
B. Discussion	47
CHAPTER V : CONCLUSION & SUGGESTION	53
C. Conclusion	
D. Suggestion	
REFERENCES	55 58

LIST OF TABLES

Table		Page
4.1	Result of Analysis Types and Tokens in the Textbooks	38
4.2	Noun Categories	41
4.3	Verb Categories	42
4.4	Adjective Categories	44
4.5	Adverb Categories	44
4.6	Pronoun Categories	45
4.7	Preposition, Conjunction, Interjection, and determiner Categories	46

LIST OF FIGURES

Figure		Page
3.1	Overview of Vocabulary Profile Tool	31
3.2	Overview of File Viewer and Editor Tool	32
3.3	An Example of A txt. File Converted	33
3.4	An Interface of AntWordProfiler	34
3.5	Press START	35
3.6	Result of the AntWordProfiler	35
4.1	Part of Speech Categories	40

LIST OF APPENDICES

Appendix I The Result from AntWordProfiler



LIST OF ABBREVIATIONS

ELT English Language Teaching

EFL English as Foreign Language

GSL General Service Lists

HFW High-Frequency Words

GE General English

CHAPTER I

INTRODUCTION

This section explains the general description of why conducting this research. It contains the background of the study, research question, objectives of the study, significance of the study, scope and limitation of the study, and definition of key terms. All of those components are required before organizing the data as the introductory chapter of this study.

A. Background of the Study

Corpus linguistics, also recognized as corpus-based studies, is a language study using specialized tools to analyze naturally occurring language data. It aims to examine linguistic elements of a specific language or language variety using empirical evidence contained in a corpus (or corpora in the Latin plural), a compilation of texts collected using well-established linguistic criteria. Then, there has been an increase in corpus-based studies and English Language Teaching (ELT) research over the last ten years.

The research on teachers and ELT scholars has recognized the importance of using didactic materials that spontaneously generate language by native English speakers (language corpora). Additionally, the authentic language generated by English learners worldwide (learner corpora) has been

¹ McKenery, T. & Wilson, A. (2001) Corpus Linguistics. Edinburgh: Edinburgh UK

² Sardinha, T.B. (2004). *Linguística de corpus*. Barueri, SP: Manole.

scrutinized to gain a more in-depth understanding of linguistic features. As a result, it is suitable for both teaching and research.

A growing number of researchers have used corpus-based approaches to study various aspects of linguistics, including the development of language teaching methods.^{3,4} Corpus linguistics has almost become a common practice in the development of dictionaries,⁵ and more recently, in the development of reference grammars.⁶ Corpora have shown to be a valuable tool for detecting non-native language qualities in the language development of foreign language learners, such as the over-representation of unique phrases or linguistic characteristics.⁷ It has questioned many conventional approaches to language studies in several ways as a pedagogical method⁸. Then, it has been shown that corpus linguistics can be effective in the research of collocational patterns and word frequency.

Textbooks are considered one of the essential components in supporting the existing curriculum being implemented. Based on Richards, a textbook is one of the instructional tools used to facilitate learning by increasing

-

³ Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208.

⁴ Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora*, second language acquisition and foreign language teaching (pp. 3–33). Philadelphia: John Benjamins.

⁵ McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574.

⁶ Burton, G. (2012). Corpora and coursebooks: destined to be strangers forever? *Corpora*, 7(1), 91–108.

⁷ Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora*, second language acquisition and foreign language teaching (pp. 3–33). Philadelphia: John Benjamins.

⁸ Hunston, S., & Francis, G. (1999). Pattern grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.

cognitive processes and providing structure and development for students to follow. 9 In general, textbooks serve two essential pedagogical functions: a curricular component that advances students' knowledge of a specific topic and a conceptual component representing the learner's cognitive constructs. 10 Thus, it indicates that textbooks are essential resources for students in the teaching and learning process.

Several studies focused on a corpus-based analysis on vocabulary in textbooks from EFL (English as Foreign Language) contexts, such as Sun & Dang in China;¹¹ Alsaif & Milton in Saudi Arabia;¹² and Nguyen in Vietnam.¹³ Only a small amount of corpus linguistic study has concentrated on textbook development and vocabulary content from a pedagogical perspective. Besides, the few research that has used corpus-based approaches to teaching materials, such as research conducted by Jamalzadeh and Chalak,¹⁴ Pathan, et al.,¹⁵ have mostly focused on academic written

⁹ Richards, J. C. (2001). *Curriculum development in language teaching*. Ernst Klett Sprachen.

¹⁰ Dharma, Y. P., & Aristo, T. J. V. (2018). An analysis of English textbook relevance to the 2013 English curriculum. *Journal of English Educational Study (JEES)*, 1(1), 24-33.

¹¹ Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93, 1-13.

¹² Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, 40(1), 21-33.

¹³ Nguyen, C.-D. (2020). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal*, 1-10.

¹⁴ Jamalzadeh, M., & Chalak, A. (2019). A Corpus-based study of academic vocabulary in physiotherapy research articles. *Language Teaching Research Quarterly*, *9*(1), 69-82.

¹⁵ Pathan, H., Memon, R. A., Memon, S., Shah, S. W. A., & Magsi, A. (2018). Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups. *International Journal of English Linguistics*, 8(4).

materials.^{16,17} Studies discussing the needs of young learners have been left unnoticed.^{18,19} Considering the significance of textbooks in foreign language learning and teaching, especially for younger students.^{20,21} So, it needs more in-depth corpus linguistic research that not only focuses on the lexis of textbooks.

Some previous studies were focused on an attempt to examine the corpus analysis of English textbooks. For example, Shin and Chon had conducted a corpus study about English textbooks in the Korean EFL context. The findings showed that secondary school textbook word lists included up to 7,430 additional words beyond the 3,000 allowed by the National Curriculum for high school.²² Furthermore, Hajiyeva conducted research that focused on corpus analysis of university textbooks on specific subjects for English majors. The finding showed that 127 academic word families were commonly used in the overall university textbook corpus to be used as part of

¹⁶ Hoey, M. (2000). The hidden lexical clues of textual organization: A preliminary investigation into an unusual text from a corpus perspective. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 31–42). New York: Peter Lang.

¹⁷ Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91–102). New York, NY: Peter Lang.

¹⁸ Foster, J., & Mackie, C. (2013). Lexical analysis of the Dr. Seuss Corpus. *Concordia Working Papers in Applied Linguistics*, 4, 1–21.

¹⁹ Keck, C.M. (2004). Corpus linguistics and language teaching research: Bridging the gap. *Language Teaching Research*, 8(1), 83–109.

²⁰ Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. *The Language Learning Journal*, 40(1), 35–45.

²¹ Skolverket. (2006). The role of teaching materials in teaching: Secondary school teachers' choice, use and evaluation of teaching materials in Arts, English and Social Sciences. Stockholm: Skolverket.

²² Shin, D. & Chon, Y.V. (2011). A corpus-based analysis of curriculum-based elementary and secondary English textbooks. *Multimedia-Assisted Language Learning*, 14(1), 149-175.

the second-year English major's word list.²³ Besides that, Hsu conducted a corpus analysis on general-purpose college English textbooks. It is primarily a corpus analysis of General English (GE) reading textbooks used in Taiwanese universities. More than 2000 words, the General English textbook can provide 162–2,001 new word families for students.²⁴ To conclude, the studies mentioned above only focus on a corpus study about English textbooks containing the word lists of textbooks.

English teacher forum is an English teacher organization to share information about the teaching and learning process. The teachers usually discuss the program that focuses on curriculum development, makes lesson plans, creates textbooks, and a fun learning process using various learning models according to each school's conditions. Therefore, the forum has been created to enhance teaching and learning qualities.

In the Tuban regency, one of the products created by the English teacher forum of Junior High School is English textbooks. The name of the textbook is "Better@English." It consists of five books that are two books for teaching grade VII, two books for VIII, and a book for XI. Based on a preliminary study, this forum had constructed the textbooks and used it for learning English in public schools for more than ten years in Tuban. On the other hand, because the researcher hopes that this result of this study might

.

²³ Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. Ampersand, 2, 136–144.

²⁴ Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage, *Electronic Journal of Foreign Language Teaching*, 6(1), 42–62.

be useful for the teachers in Tuban as the suggestion or consideration in designing the appropriate textbook especially English textbooks.

Before constructing the textbook for application in the following school year, the English teacher forum always improves the quality of books regularly every year based on evaluations from the teachers who have used this book to teach and learn in the classroom. The review is carried out to protect the textbooks' quality and relevance based on the applicable curriculum. Subsequently, before constructing the English textbooks, the forum discusses the material that is appropriated for teaching the students in Junior High School based on grades VII, VII, and XI. Then, they construct, write, and publish the textbooks to facilitate the Junior High School teachers and students in Tuban through some rich learning materials for the teaching and learning process.

This present research focuses on a corpus analysis of the English vocabulary that is used by English teachers to construct English textbooks in Indonesia, especially in the EFL context. This study discusses a corpus study of English textbooks constructed by the English teacher forum of Junior High School in Tuban. The previous studies' differences include a corpus study that investigates the English vocabulary that English teachers frequently use, especially in Indonesia. On the other hand, no corpus study examines textbooks from Indonesia. Moreover, there is some research about corpus studies, but it does not specifically discuss the textbooks. Furthermore, the research was carried out to develop knowledge about the English vocabulary

that Indonesian teachers frequently used to construct English textbooks. This will enrich the knowledge on the varieties and features of English words used in the EFL context to further develop applied linguistics theories.

B. Research Question

According to the background of the study, the following is the research question for this study:

What are the corpora of English vocabulary frequently used by the English teacher forum of Junior High School to construct the textbook in Tuban?

C. Objectives of the Study

Based on the formulation of the research question, this study investigates the corpora of English vocabulary that is frequently used by the English teacher forum of Junior High School to construct the textbook in Tuban.

D. Significance of the Study

The findings of this study have the potential to help the teacher, the reader, and the future researcher.

1. For the teacher

This study provides teachers with useful information about vocabulary coverage characteristics in their textbooks and materials

to meet the appropriate students' level of learning. It also emphasizes teachers' importance in considering several factors when adjusting the vocabulary level with students' background experience before selecting or constructing a textbook based on students' needs.

2. For the reader

This research may give a detailed description of vocabulary used in a Junior High School textbook, enrich the knowledge about the English vocabulary used frequently by English teachers in Indonesia, and how to use a vocabulary counting application.

3. For the future researcher

By this result, the researcher hopes that the future researchers who conduct the same topic can enrich their knowledge for the following research and use this study as a reference to help them with the corpus study. It will be preferable to include additional sources, such as spoken corpora, to determine if the developed wordlist responds to genre changes in the same discipline.

E. Scope and Limit of the Study

The scope of this research focuses on the English textbooks for Junior High School that have been written and constructed by the English teacher forum in Tuban. Five books are constructed to teach junior high school grades VII, VII, and XI in Tuban. The title of these textbooks is

Better@English. This research analyzes the English textbooks constructed for the academic year 2019/2020.

This research's limitation analyzes the English vocabulary of textbooks that has been constructed by the English teacher forum in Tuban for Junior High School. All of the vocabulary items in these textbooks are analyzed except the figures, tables, footnotes, cover, references, and appendices. It means that the research analyses how many and how often the teacher uses that vocabulary. Then, the researcher analyses more deeply the English vocabulary in lexical analysis and lemmatization. Finally, the researcher computes all the vocabulary in the textbook using AntWordProfiler 1.4.0 software by Laurence Anthony from Waseda University to generate a list of lexical items used in the textbook.

F. Definition of Key Terms

To avoid readers misinterpreting the objective of this study, the definition of key terms can aid in clarifying the interpretation of various elements in this research.

1. Corpus

A corpus (plural Corpora) is a collection of linguistic data typically stored as an electronic archive and collected as written texts or transcriptions of spoken language. Furthermore, the computer corpora can contain millions of running terms, the characteristics of which can

be analyzed using tagging and concordance programs.²⁵ In this study, a corpus is described as a set of written vocabularies that can be used to standardize vocabulary measurement in the English textbook created by the Tuban English teacher forum that "Better@English" is the title of the book.

2. Vocabulary

Vocabulary is one of the language components that people are expected to acquire while studying a language. It is expected to be the primary focus of foreign language study because there is a possibility that learning a foreign language is close to comprehending its vocabulary. The significance of vocabulary in the teaching and learning of English cannot be overstated. It is regarded as the foundation for acquiring knowledge of a second language.²⁶ In this research, vocabulary is defined as the words that appear in the textbook developed by the English teacher forum in Tuban.

3. English Textbook

One of the types of instructional materials used in the study of English is an English textbook. It is just one of many possible learning materials that can assist teachers and students in achieving their learning objectives. The textbook is usually written in a concise, well-structured,

_

²⁵ What Is a Corpus, What Is Corpus Linguistics?" (n.d.): 3.

²⁶ Lotfi, Ghazal, Learning Vocabulary in EFL Context Through Vocabulary Learning Service, Novitas Royal, Vol. 1, no. 2, 2007

and greatly simplified manner.²⁷ An English textbook is defined in this study as a supplement material for Junior High School English called "Better@English," which was created by the English teacher forum in Tuban.



_

 $^{^{27}}$ John W, Cresswel. Qualitative Inquiry And Study Design: Choosing Among Five Approaches Second Edition (New Delhi. 2007),156

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter describes the theory related to the topic of this research and previous study. The theoretical framework discusses corpus and corpus analysis, vocabulary, textbooks, and the use of corpora in English Language Teaching (ELT). Besides, some previous studies related to corpus analysis are revealed.

A. Theoretical Framework

The researcher uses some theories to support the current study, which is described in detail below.

1. Corpus and Corpus Analysis

a. Definition of Corpus

Corpus (plural Corpora) is a term that refers to a grouping of written or spoken words that have been saved on a computer.²⁸ The software that is specifically designed can collect and analyze a large number of phrases. Then, a corpus is a collection of written or spoken texts as defined by the Oxford Advanced Learner's Dictionary.²⁹ In the current studies, a corpus is commonly defined as a collection of authentic machine-readable text chosen to characterize or represent a state or variety of a language.³⁰ Similarly, Linguists agreed that corpus is related to the electronic corpus.³¹ He defined corpus as a collection of texts stored on digital mediums for use by linguists in retrieving linguistic items for research or by lexicographers in creating dictionaries.³² Based on the definitions above, it is possible to conclude that corpus is stored in a computer-based system and contains complete

²⁸ Anne O'Keeffe, Michael Mccarthy, and Ronald Carter, "From Corpus to Classroom: Language Use and Language Teaching" (n.d.): 1.

²⁹ Oxford Advanced Learner's Dictionary (9th edition): Oxford University Press

³⁰ Grigaliuniene, Jone. 2013. Corpora in Language Studies. Lituania: Vilnius University.

³¹ Lindquist, H. 2009. Corpus Linguistics and the Description of English. Edinburgh: Edinburgh University Press.

³² Ibid,.

forms of text, whether written or spoken, and it may be claimed that computers play a key part in the corpus.

b. Types of Corpora

A corpus cannot be easily classified into a specific category. On the other hand, the corpus may have characteristics or properties for grouping. Therefore, there are some types of corpus categorized below.³³

1) Monolingual Corpus

The most common type of corpus is a monolingual corpus. This is a corpus that only contains texts in one language. The corpus is usually annotated with tags for various parts of speech. It is used by a diverse range of users for various tasks ranging from the most practical, such as verifying proper word usage or looking up the most natural word combinations, to the scientific, such as finding common patterns or new trends in language. For example, Tehran Monolingual Corpus (TMC).

2) Parallel corpus

At least two monolingual corpora compose a parallel corpus. These corpora are translations of each other. For example, a novel and its translation, or the translation memory of a CAT tool, could be used to generate a parallel corpus. Both languages must be aligned, which entails matching corresponding segments, which are usually sentences or paragraphs. Then, the user can conduct a search in one language for all instances of a word or phrase, with the results displayed alongside the corresponding sentences in the other language. The user can then view the translation of the search word or phrase. For example, Tehran English-Persian Parallel Corpus (TEP).

-

³³ https://www.sketchengine.eu/corpora-and-languages/corpus-types/

3) Comparable corpus

A comparable corpus is one of two or more monolingual corpora, typically in different languages that was built using the same principles. As a result, the content is similar, and results can be compared across corpora even though they are not translations of each other (and therefore, they are not aligned). Moreover, when users search these corpora, they can take advantage of the fact that the corpora also have the same metadata. The examples of the comparable corpus are Araneum corpora and the International Corpus of English (ICE). Besides, Comparable corpora in Sketch Engine is CHILDES corpora or various corpora made from Wikipedia.

4) Diachronic corpus

A diachronic corpus is a corpus that contains texts from various periods and is used to study the evolution or change of language. Sketch Engine allows you to search the corpus as a whole or just a specific time interval. Besides, "Trends" is a specialized diachronic feature that identifies words whose usage changes the most over the specified time.

5) Synchronic corpus

A synchronic corpus is the opposite of a diachronic corpus, as it contains texts from the same point in time. It is a snapshot of language at a single point in time. The "enTenTen" corpora are such snapshots because their content is gathered in a matter of months.

6) Learner corpus

A learner corpus is a set of texts written by language students. The corpus is being used to investigate the errors and challenges that the learner faces when learning a new language. The Louvain Corpus of Native English Essays (LOCNEE) and the International

Corpus of Learner English (ICLE) are two examples of the learner corpus.

7) Multimedia corpus

A multimedia corpus is a group of texts with audio, video, or other types of multimedia content added to it. For example, in Sketch Engine, the British National Corpus (BNS) spoken component offers links to matching recordings that may be listened to straight from the interface. Additional corpora may include videos in which the corpus text is spoken and images of the original manuscript or printed copy of the text.

8) Specialized corpus

A specialized corpus is a group of texts that are limited to one or more subject areas, domains, or topics. This type of corpus is used to investigate how a specific language is used. Sketch Engine, for example, can create specialized sub corpora from general corpora such as the Cambridge and Nottingham Corpus of Discourse in English (CANCODE).

This present research discusses the monolingual corpus because it collects only English vocabularies that the English Teacher Forum frequently used in constructing textbooks. Besides, a synchronic corpus is discussed from the vocabulary of the textbooks that are constructed at the same point in time for the 2019/2020 academic year.

c. Corpus Analysis

According to Kaewphannagam et al., corpus-based analysis has been used in language teaching and learning to develop and evaluate vocabulary for a long time.³⁴ In other words, corpus analysis is a sort of language analysis that occurs naturally through the use of corpora and computerized programs. Thus, the corpus analysis is carried out with

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

³⁴ Kaewphannagam, C., Broughton, M.M., & Soranasataporn, S. 2002. Corpus-Based Analysis: Guidelines for Getting Practical Language Input in Materials Development.

computer-based programs' assistance; for example, corpus-specialized software is used to generate the frequency of words.

McCarthy states that someone can learn about the most common words and phrases in English by using corpus analysis. For example, the distinctions between written and spoken English, the most common tenses used, what prepositions come after specific verbs, how people use words like *can*, *may*, *and might*, the distinctions between formal and informal words, and how and why they are used. Furthermore, we can see how language is used in modern society, how language has changed over time, and how language is used in various situations by applying the concept of corpora. Thus, a corpus analysis enables the researcher to learn about the various ways in which languages are used. In another way, corpus analysis is empirical research because it examines actual usage patterns in natural texts.

The corpus analysis results are typically displayed in two ways: concordance figures or a word frequency list. Then, the tools of corpus analysis are used to calculate those results. The following are some of the fundamental techniques or tools that can be used in corpus analysis.

1) Collocation

Bennett defines collocation as the statistical tendency of words occurring together.³⁵ It means that if one word is used, there is a high statistical probability that another word or series of words will follow. For example, *drop off, swimming pool, letterbox, etc.* We can study collocation using computerized corpora by observing concordance lines to see which words occur to the right or the left of the search word. Collocation can also be used to help us understand specific words used in a phrase.

2) Frequency Lists

.

³⁵ Bennett, Gena R. 2010. Using Corpora in the Language Learning Classroom. Michigan ELT.

According to Grigaliuniene, a corpus-based frequency list is a list of all the terms that appear in a corpus.³⁶ Even though frequency data cannot be used as the sole source of information for a learner when deciding which words to learn, prioritizing which words must be learned first will be a good starting point. At its most basic, frequency refers to the number of occurrences in a corpus (or text). It can obtain rank ordering of all the words according to frequency by executing a word frequency list. Furthermore, the frequency list can be sorted by overall frequency or alphabetically, identifying the most and least common words or phrases used in a corpus. As a result, it can help the researcher select the most frequent or least frequent lexical items in language research.

3) Concordance

Concordance displays a list of how words and phrases appear in the context of the specific text. Concordance can be used to find the examples of specific words used in text faster than searching the entire text. Furthermore, Grigaliuniene defines concordancer as a type of search engine designed for language study. It examines a corpus and records every occurrence of a word or phrase. Concordance is used by entering a word or phrase, and the software then searches for all examples of that word or phrase. In a matter of seconds, a computer can search through millions of words. Then, the keyword is frequently referred to as the "node," and concordance lines are typically displayed with the node word or words in the center and seven or eight words on either side, a technique Key-Word-In-Context (KWIC) known concordance.

³⁶ Grigaliuniene, Jone. 2013. Corpora in Language Studies. Lituania: Vilnius University.

2. Vocabulary

a. Definition of Vocabulary

The Cambridge Advanced Learners' Dictionary gives a common definition of vocabulary as "all the words that exist in a particular language or subject." Several experts defined vocabulary terminology in various ways. According to Hornby, vocabulary includes all of the known and used words in a given language. While Richard and Renandya identified vocabulary as a fundamental component of competency, it also serves as the foundation for effectively committing to four language skills: speaking, listening, reading, and writing. Other scholars, such as Lehr, Osborn, and Hiebert, defined vocabulary as knowledge of words or their meanings expressed in spoken, written, productive or receptive language. According to some experts' definitions of vocabulary terms, it refers to all of the words that people use in any language, whether written, spoken, or otherwise.

Vocabulary is an essential component of English language learning. Teachers will be unable to deliver or express their ideas about the material without it, and students will be unable to understand what the teacher's saying. To ensure that our students are satisfied with our explanations, we should continue to improve our ability to acquire and use appropriate vocabulary when delivering the material as a good teacher. As a result, a learner's vocabulary can be a good indicator of how well they understand the language.

³⁷ www.cambridge.org,

³⁸ Budi Setiawan, 'Improving The Students Vocabulary Mastery through Direct Instruction' (Universitas Negeri Surakarta., 2010).

³⁹ Zahrotul Izzah, 'The Use of Word Chain Game to Improve Vocabulary Mastery Of Eight Grade Students at SMPN 3 Kalibagor' (Universitas Muhammadiyah Purwokerto, 2015).

⁴⁰ Christine Dugan, *Strategies for Building Vocabulary in Language Arts* (Huntington beach: Shell Education publishing Inc, 2010).

b. Types of Vocabulary

Differentiating vocabulary types can assist teachers in determining how much time and effort they should commit to teaching specific words. Nation divided vocabularies into four categories based on how frequently they appeared in texts.⁴¹

1) High-Frequency Words (HFW)

The term "High-Frequency Words" refers to words that frequently appear in any text written in any language. Among these words are frequently found function words such as articles, prepositions, and some content words. Additionally, Nation referred to Michael West's General Service Lists (henceforth GSL)⁴², which contained around 2000 words families, as the most popular HFW. ⁴³

2) Academic Vocabulary

Academic vocabulary consists of words that are commonly used in academic discourse or texts. In this case, Coxhead's Academic Word List/AWL is considered the academic vocabulary list, and it contains 570-word families that are not included in the GSL but frequently occur in a corpus of academic texts from various disciplines.

3) Technical Vocabulary

Technical vocabulary is so specific to texts in a particular subject area that it is uncommon to come across it in other texts. Furthermore, the specific technical word list items varied by discipline, interest, and specialization. Technical words include those compiled in specialized dictionaries to meet the needs of English as a second language (ESL) or English as a foreign

⁴¹ Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

⁴² West, M. (1953). A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology. London: Longman, Green & Co.

⁴³ Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

language (EFL) students studying a particular topic at Englishspeaking universities and non-English speaking practitioners of a particular occupation.

4) Low-Frequency Words

Low-frequency words infrequently appear and in small proportions compared to high frequency and academic words. The words classified as low frequency or high frequency varied by discipline and genre/text type. Low-frequency words included archaic words from an older dialect or variation of a language, infrequent words due to social norms, words from other languages, and proper nouns (names of people and places).

There are two methods for determining vocabulary. The first method is to count the number of tokens (which is how coverage vocabulary is calculated), and the second method is to count the number of types, lemmas, or word families. As a result, understanding the various methods for counting words is critical. There are several ways to determine what will be counted when counting words, according to Nation.⁴⁴

1) Tokens

The total number of words is referred to as tokens. Even if the same word form appears numerous times in the phrase, it will be tallied. For example, the statement "it will be difficult to pronounce it correctly" comprises eight tokens.

2) Types

If a word appears more than once, it will not be counted again using this method of counting words. As a result, there are seven types of sentences in the eight-token sentence "it will be difficult to pronounce it correctly."

3) Lemmas

.

⁴⁴ Nation, P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

In each word class, lemmas are the root forms. It consists of a headword and various inflected forms of it. The plural, past tense, present tense, present participle, past participle, comparative, superlative, and possessive inflections are all used in English. For example, the verb: "jump, jumps, jumped, jumping," the noun: "jump, jumps", the noun: "jumper, jumpers" are 3 lemmas, 6 types, 8 tokens.

4) Word family

The root forms of words are word families. There is a root form shared by all regularly inflected and derived words. The base word's meaning must be closely related to the derived word's meaning. For example, "kick, kicks, kicked, kicking, kicker, kickers" and "do, doing, undo, redo, doable" are consist of 2-word families, 7 lemmas, 11 types, and 11 tokens.

Then, in order to measure vocabulary profile by AntWordProfiler, token percentages are used as a method of counting words in this research.

3. The Use of Corpora in English Language Teaching (ELT)

Corpora have introduced new language study approaches and changed the perspectives of linguists in fields such as lexicography and English Language Teaching (ELT). As explained below, several aspects of language pedagogy in ELT demonstrate how corpus findings have been put to practical use.

a. To Compile English Dictionaries and Teaching Materials.

If one wishes to publish a dictionary that accurately reflects how English is spoken daily, accurate language information must be obtained. The only way to do so is to look into how native English speakers use the language. Corpora provide a trustworthy source of authentic language data, and as such, they can be used to create a variety of pedagogical materials such as dictionaries, course books, and

vocabulary books. For example, in the past, materials developers relied on their intuition when deciding on the content of course books. Even native speakers, however, are inaccurate in estimating the frequency of use of various linguistic elements. As a result of the availability of large databases of language data, lexicographers and materials developers search corpora, and their findings influence what is included in dictionaries and course books.

b. Reference Tools

When English teachers or students are unsure whether certain grammar and lexis elements are correct, they can look them up in corpora. This is especially important for non-native English teachers, who are frequently asked by their students why certain English words are used in the way they are. Even if non-native teachers speak English fluently, their second language and their intuitions may be suspect. As a result, teachers can consult corpus data whenever there is a doubt to ensure that what they teach is correct.

c. Data-Driven Learning (DDL)

According to this methodology, the language learner "is also a research worker whose learning must be driven by access to linguistics data." Students independently identify common patterns in lexis and grammar by exploring authentic language material, with the teacher merely facilitating the entire process. This inductive approach enables students to discover which linguistic forms are used in communicative contexts and increases their understanding of how language works in everyday situations.

d. Large Databases of Language

Corpora are large collections of languages generated by students. This allows researchers to examine learners' linguistic

development at various proficiency levels and gain valuable insight into which aspects of English cause second language learners' problems. For example, even at the advanced level of proficiency, students misuse or underuse vocabulary, giving the impression of 'non-nativeness' of learner English when compared to native speakers' performance. The ICLE corpus, as mentioned above, is an interesting example. It is also possible to create corpora that tracks a group of people's linguistic development over a longer time. It's difficult and time-consuming to collect data like this, but longitudinal studies like these help us understand how a second language develops over time.

e. Affect the Classroom-Based Research

It is self-evident that pedagogically-oriented research aims to investigate classroom processes and, as a result, improve how learners are assisted on their path to language proficiency. Before any teaching implications can be formulated, researchers must have access to a representative sample of language data so that corpora can be incredibly helpful when a researcher wants to focus on specific language features.⁴⁵

4. Textbook

a. Definition of Textbook

In English language teaching, the terms "course books" and "textbooks" are used interchangeably. Nonetheless, some EFL experts define textbooks as a collection of instructional materials used in the classroom to teach English as a subject. They are also specifically designed for specific target audiences in a specific setting. Textbooks, for example, are widely used across the country in a specific educational context. While textbooks are materials used in the

-

⁴⁵ Szudarski, P. (2011). The Role of Language Corpora in Teaching English as a Foreign Language in Poland.

classroom to teach English as a subject in a specific educational setting, course books are designed for broad target audiences in both English and non-English speaking countries. Thus, a textbook is an instructional material of English as a subject taught in schools that is designed for specific groups of learners in a defined context, such as a specific educational context within a specific country.

ELT practitioners have defined textbooks in a variety of ways. They are described as a learning tool,⁴⁶ a principle,⁴⁷ and a staple⁴⁸ in many ESL/ EFL classrooms. Textbooks serve as a learning aid that makes it easier to learn. Furthermore, they are useful in teaching and learning for both teachers and students. Finally, textbooks serve as a guide used systematically in the classroom as the foundation of a language course in their application as a principle.⁴⁹ To sum up, textbooks are supplementary materials that play a critical role in the teaching and learning process.

B. Previous Studies

Some previous studies discuss corpus studies related to investigating the vocabulary.

First, the research title is "A Corpus-based study of Academic Vocabulary in Physiotherapy Research Articles" by Mehri Jamalzadeh and Azize Chalak.⁵⁰ The distribution of high-frequency words in Coxhead's Academic Word List (AWL) and West's General Service List (GSL) was compared to a list of academic words often occurring in physiotherapy research publications in this quantitative study. According

⁴⁶ Graves, K. (2003). Coursebooks. In Nunan, D. Practical English language teaching. New York: McGraw Hill Inc.

⁴⁷ Ur, P. (1996). *A course in language teaching : practice and theo* ry. Cambridge : Cambridge University Press.

⁴⁸ Garinger, D. (2001). Textbook evaluation. TEFL Web Journal.

⁴⁹ Ur, P. (1996). A course in language teaching: practice and theory. Cambridge University Press.

⁵⁰Jamalzadeh, M., & Chalak, A. (2019). A Corpus-based study of academic vocabulary in physiotherapy research articles. *Language Teaching Research Quarterly*, 9(1), 69-82.

to the results, AWL contains 570-word families, 562 of which frequently occur in physiotherapy research articles (RAs), covering 11.51 % of the tokens in the corpus. Additionally, the Academic Word List (AWL) provides benefits for physiotherapy students due to limited coverage for certain word families. Besides, there is a lack of frequently used physiotherapy academic words.

Second, the study was conducted by Ping Wang. The title of the research is "A Corpus-based Study of English Vocabulary in Art Research Articles." The purpose of this study was to enhance the effectiveness of art majors' English learning by conducting a quantitative analysis of high-frequency words found in art research articles, particularly in academic reading and writing. The result showed 2242 words families with 95% families with 95% of lexical coverage in the Art Research Article Corpus (ARAC). Additionally, both the General Service List (GSL) and the Academic Word List (AWL) had a high coverage level in their corpora. In contrast, the Art Research Article Corpus (ARAC) had a different order of high-frequency words.

The third previous study came from Ismail Xodabande and Nasrin Xodabande. The research title is "Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study." This study aimed to examine the lexical characteristics of psychology research articles using both West's General Service List (GSL) and Coxhead's Academic Word List (AWL). They used the computer software AntWordProfiler to analyze the data. To that end, they examined a corpus of approximately 74 million words comprising 8,500 psychology research articles. According to the results, 1,537 high-frequency AWL and non-GSL/AWL word types (rather than word families) comprised 17.91 % of the corpus, while the top 570-word types covered 13.44 % of

.

⁵¹ Wang, P. (2017). A corpus-based study of English vocabulary in art research articles. *Journal of Arts and Humanities*, 6(8), 47-53

^{52 &}quot;Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study

the corpus, which is more than the 570 AWL word families combined (with about 3000 types).

The fourth previous study was conducted by Habibullah Pathan, Rafique A. Memon, Shumaila Memon, Syed Waqar Ali Shah, and Aziz Magsi. The research title is "Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups." In this study, they examined how often AWL Word Families appeared in Pakistani doctoral thesis texts. The data were derived from 200 doctoral theses in two major scientific disciplines. They use concordance software AntConc by Anthony Laurence. According to the study, AWL words cover 8.76 % (496839 words) of Pakistani doctoral theses. They attempted to extend the analysis of the AWL text coverage by taking into account the frequency with which word families occur. Based on the result, out of 570 AWL word families, 550-word families with a total of 96.49 % occurred more than ten times in Pakistani doctoral theses.

The fifth previous research was conducted by Farrokh Heidari, Alireza Jalilifar, and Anayatollah Salimi. The research title is "Developing a Corpus-Based Word List in Pharmacy Research Articles: A Focus on Academic Culture." The development of a Pharmacy Academic Word List (PAWL) is the subject of this study. According to the findings, the created PAWL has 750-word families that make up 17.69% of the corpus under investigation. The study's outcomes highlight the need to create domain-specific academic word lists to address the demands of non-native researchers and postgraduate students in a variety of disciplines. This word list can be used to create an EAP lexical syllabus as a starting point.

-

⁵³ Pathan, H., Memon, R. A., Memon, S., Shah, S. W. A., & Magsi, A. (2018). Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups. *International Journal of English Linguistics*, 8(4).
⁵⁴ Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a Corpus-Based Word List in Pharmacy Research Articles: A Focus on Academic Culture. *International Journal of Society*, *Culture & Language*, 8(1), 1-15.

The sixth previous research was "Accounting Academic Word List (AAWL): A Corpus-Based Study" by Reza Khany and Behrooz Kalantari. The aim of this research is to develop a field-specific academic word list for accounting, to find the degree of coincidence between the word list and Coxhead's academic word list (AWL), and also to compare the occurrences of the most frequently used words in the list with six available word lists in different disciplines. Range software was used to compile and analyze a large corpus of accounting research articles. Further investigation revealed that only 354 of the identified high-frequency word families corresponded to those listed in AWL. Furthermore, the list's top 50 most frequently used words accounted for 3.98% of the total corpus.

Based on previous studies that showed before, all researchers are implementing a corpus study to analyze and describe how many academic vocabularies are frequently used in the specific discipline. Therefore, the focus of this study also analyzes the vocabulary in EFL contexts using corpus analysis. However, this research focuses on not only the academic vocabulary but also all of the English vocabularies that teachers used to construct English textbooks.

CHAPTER III

RESEARCH METHOD

This chapter presents the research method used in this study. It consists of research design, research setting, research subject, data and source of data, research instrument, data collection technique, and data analysis technique.

A. Research Design

A quantitative descriptive method is the design of this research. The design is chosen based on the fact that it has been widely used in most corpus approach studies. Based on Creswell, quantitative is gathering numerical data using an instrument from many people who have previously determined questions and responses.⁵⁵ It emphasizes that quantitative research considers exploring numeric data to collect and analyze the result of the study. It means the method to discover the limit of optimal frequency for vocabularies.

This research investigates a corpus study of vocabulary in teacher-constructed textbooks by the English teacher forum of Junior High School in Tuban. The fact of the analysis was revealed in depth. These corpus studies represent data analysis in the form of words, phrases, or sentences collected from the textbook. On the other hand, using a corpus study enables more rapid and precise analysis of usage frequency, which is the primary focus of corpus linguistics analysis. Besides, by utilizing a corpus and focusing on the frequency of linguistic units, it is possible to ascertain what is truly

⁵⁵Creswell, J.W. (2012). Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, 4th ed. Boston: Pearson

significant. According to Baker, frequency indicates significance; it implies that something that occurs frequently is a significant concept. In terms of linguistic analysis, a linguistic unit that occurs frequently in a text denotes the unit as a text's characteristic feature. Therefore, the selection of quantitative design is suitable for the aim of the research. Thus, the quantitative method is used for this study.

B. Subject and Setting of Study

This research's subject is five books constructed by the English teacher forum in Tuban for teaching students of Junior High School grades VII, VII, and XI in Tuban. The title of the textbook is "Better@English" for the academic year 2019/2020. All of the vocabularies in these textbooks are analyzed except the figures, tables, footnotes, cover, references, and appendices. The setting of this study is on Tuban regency, East Java.

C. Data and Source of Data

1. Data

This research is conducted to investigate the English vocabulary frequently used by the English teacher forum of Junior High School to construct the textbook in Tuban. For this study's aim, the data gathered was English vocabulary that teachers used in the constructed English textbook.

2. Source of Data

The document needs to be analyzed because the researcher is curious to recognize the English vocabulary that the English teachers frequently used to construct the textbook. The source of data in this research is the English textbooks that are entitled "Better@English." The English Teacher Forum published this English textbook in Tuban for the academic year 2019/2020.

D. Research Instrument

An instrument is a tool that aids in the data collection process. The researcher was also a key instrument in this study because she collected and processed data simultaneously. On the other hand, the researcher used software with a multiplatform tool and tagger that is frequently used in corpus-based research to facilitate the analysis of the English textbook in this study. It is possible to assess the frequency counts of language in large numbers using corpus-based tools, which would be impossible to do manually. To help the researcher organize the data, a user-friendly software AntWordProfiler 1.4.0 is used as another instrument. AntWordProfiler is a multiplatform freeware tool that was used when conducting corpus linguistics research on vocabulary profiling. It is an application from Anthony Laurence

from Waseda University to measure the data and to standardize this research.⁵⁶

AntWordProfiler contains two tools. The following is the explanation of the feature.

a. Vocabulary Profile Tool

The main function of this software is a tool for general vocabulary profiling. It appears in the main window of the software. For a corpus of texts loaded into the application, this tool creates vocabulary statistics and frequency data. It compares the files to a set of vocabulary level lists based on research by Paul Nation, which can be simple frequency lists or "family lists."

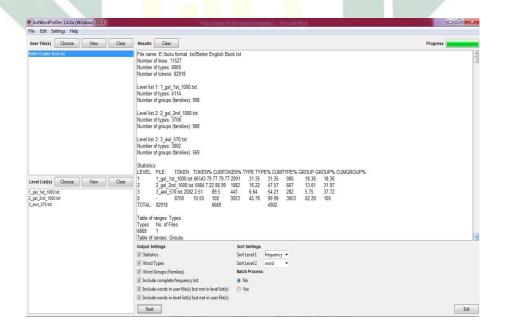


Figure 3.1 Overview of Vocabulary Profile Tool

⁵⁶ Anthony, L. (2014). AntWordProfiler v.1.4.1 [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software.

b. File Viewer and Editor Tool

This tool helps to view an individual user file and use color-coding to highlight the various levels of vocabulary in the file. It also displays the total coverage of various vocabulary levels. The tool allows editing the file and instantly sees how the changes affect the vocabulary level of the words and the text. This helps to clarify the text. When the thesaurus option in the main menu is selected, vocabulary items with thesaurus entries are highlighted, and similar items at a lower (or higher) vocabulary level can be selected as substitutes.



Figure 3.2 Overview of File Viewer and Editor Tool

E. Data Collection Technique

The purpose of the data collection technique is to describe how the researcher gathered the field's data. This research used document analysis of English textbooks that the English Teacher Forum in Tuban constructed. For

constructing the English textbooks, the forum discussed the material appropriate for teaching the students in Junior High School based on grades VII, VII, and XI. Then, they constructed, wrote, and published the textbooks to facilitate teachers and students in Tuban through some rich learning materials for the teaching and learning process.

In this research, the researcher uses the documentation method to select and gather the English textbooks for the academic year 2019/2020 from the English teacher forum in Tuban. According to Bogdan and Biklen, documentation is the process of gathering data or resources such as autobiographies, diaries, newspapers, books, articles, newsletters, and so on. Then, after the researcher obtained the textbooks, the researcher retyped the book then converted it in format plain text document (.txt) as a soft file, categorizing the data into more specific classification to make more convenient analysis and drawing on the interpretation. Furthermore, the researcher analyzed the corpora from those English textbooks.

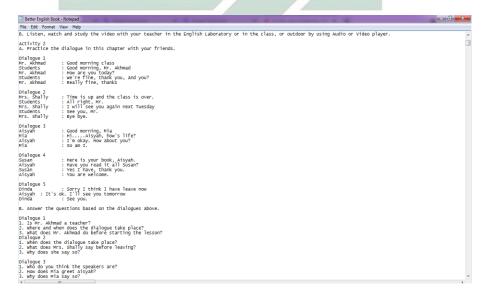


Figure 3.3 An example of a txt. file converted

F. Data Analysis Technique

Data analysis techniques were explained below.

- 1. After gathering the five English textbooks and retyping them in the format file document (.doc), then, errors and unfamiliar languages were edited or removed.
- 2. The researcher converted the document (.doc) into plain text (.txt) format.
- 3. The researcher inserted the (.txt) document into the User File box of the application.

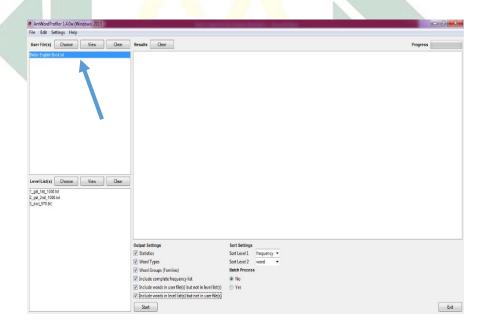
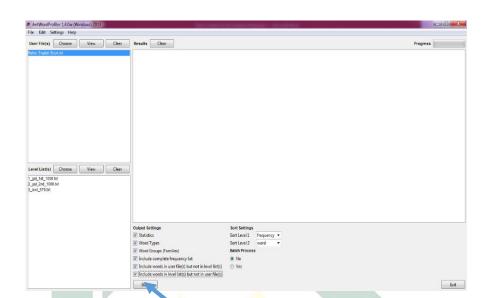


Figure 3.4 An Interface of AntWordProfiler



4. Press START and wait until generating the result of data.

Figure 3.5 Press START

5. The application displays the results of data from the document, which includes tokens, percentages of tokens, types, percentages of type groups (word families), percentages of the group (word families), and a list of examined vocabularies.

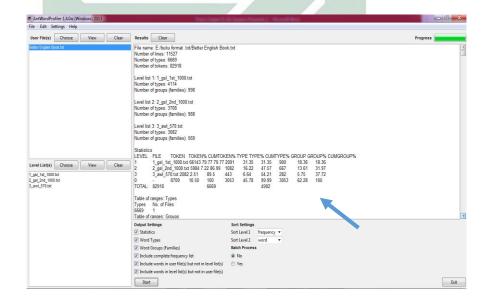


Figure 3.6 Result of the AntWordProfiler

6. Then, the result analyzed how many tokens and types to identify the amount of English vocabulary, then analyzed the lemma and limitation to have rich data from the English vocabulary that English teachers frequently used to construct the textbook.

G. Checking Validity of Finding

After analysis of the data, the research finding needs to be checked to gain a valid result and verify the trustworthiness of the data. The researcher used TagAnt 1.2.0, a simple freeware tagging tool developed by Laurence Anthony. It is used to verify that all the vocabulary that the researcher has compiled is true. After the corpus is analyzed by using AntWordProfiler, the TagAnt is used to tag the part of speech in each text's vocabulary. Then, the Tagging tool classified each word according to noun, pronoun, adverb, adjective, verb, etc. After the words were tagged using the tagging tools, the researcher sorted and analyzed all the words from the corpus of English textbooks.

CHAPTER IV

RESEARCH FINDINGS AND DISCUSSION.

This chapter presents the findings and discussion of the research conducted using the method described in Chapter III. The research finding presents the data, and the discussion shows the analysis of the research finding. The explanation about this chapter is present below.

A. Research Finding

This section addresses the answer to the research question regarding the corpora of English vocabulary that are frequently used by the English teachers' forum at the Junior High School in Tuban to construct the textbook. Five books are constructed to teach junior high school grades VII, VII, and XI in Tuban. The title of the textbook is Better@English. In this research, it analyzes the English textbooks constructed for the academic year 2019/2020. The vocabulary items in these textbooks are analyzed except the figures, tables, footnotes, cover, references, and appendices.

. The AntWordProfiler software is used to evaluate those five books. This software generates the statistical vocabulary and frequency data from the textbooks. As high-frequency words, this software identified General Service Lists (GSL) 1st 1000 and General Service Lists (GSL) 2nd 1000. Academic Word List (AWL) 3rd 570 as academic words, and "Level 0" or "groups not found (in both GSL and AWL lists)" as Low-frequency Words.

1. Types and Tokens in the Textbooks

The table below describes the total types and tokens calculated based on General Service Lists (GSL) 1st 1000, General Service Lists (GSL) 2nd 1000, Academic Word List (AWL) 3rd 570, and "Level 0" or "groups not found (in both GSL and AWL lists)." The figures are presented in Table 4.1 below.

Table 4.1 Result of Analysis Types and Tokens in the Textbooks

Level	File	Token	Token %	Cum- Token%	Туре	Type%	Cum- Type%	Group	Group %	Cum- Group%
1	GSL 1st 1000.txt	66143	79.77	<mark>7</mark> 9.77	2091	31.35	31.35	900	18.36	18.36
2	GSL2 nd 1000.txt	5984	7.22	86.99	1082	16.22	47.57	667	13.61	31.97
3	AWL 3 rd 570.txt	2082	2.51	89.5	443	6.64	54.21	282	5.75	37.72
0		8709	10.50	100	3053	45.78	99.99	3053	62.28	100
	Total	82918			<mark>6669</mark>			4902		

AntWordProfiler calculation of the corpora shows 82,918 tokens and 6,669 types from five English textbooks. The 2,000 most frequent English words listed in GSL accounted for 72,127 tokens (or 86.99 % of all running words) and 3,173-word types (47.57 % of all word types). The

AntWordProfiler had classified the words in the target corpus as belonging to the GSL base lists automatically (GSL first 1000 and GSL second 1000 most frequent English word family).

The category of academic words reported in AWL contained a total of 2082 tokens (or 2.51% of all running words) and 443-word types (or 6.64% of all running words). A total of 8709 tokens (or 10.50% of the total running words) belonged to the category of the "Level 0" or "groups not found (in both GSL and AWL lists)." Then there was "token" which was the total amount of words in the textbook's corpus. The percentages of vocabulary were token percent and type percent, with cum token percent and cum types percent representing the cumulative percentage of one vocabulary group. Meanwhile, "types" was or is referred to as an independent term.

2. Frequency of Vocabulary based on the part of speech

Based on the Got Grammar book by Ustater, noun, verb, adjective, adverb, preposition, pronoun, conjunction, interjection, and determiner are the nine parts of speech categories. Below is the explanation of vocabulary frequency based on the part of speech found in the teacher-developed textbook under study.

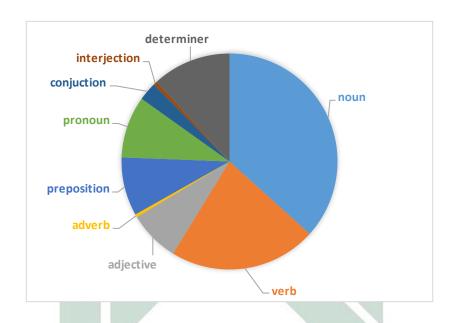


Figure 4.1 Part of Speech Categories

According to figure 4.1 above, the frequency of noun (N) 25383 lemma and the percentage is 36%, verb (V) 15262 lemma and the percentage is 21%, adjective (Adj) 5288 8%, adverb (Adv) 312 lemma and the percentage is 1%, preposition (P) 6059 lemma and the percentage is 9%. Besides, the frequency of pronoun 6403 lemma and the percentage is 9%, conjunction 1921 lemma and the percentage is 3%, interjection 376 lemma and the percentage is 1%, and for the determiner 8173 lemma and the percentage is 12%. The data shows that the teacher-developed textbooks mainly use nouns, verbs, and determiners.

Based on the validity of the result using TagAnt, the nine-parts of speech are differentiated in some categories based on the "Part of Speech" (POS) tagset that is displayed in the table below.

a. Noun Categories

The categories of the "noun" in the Part of Speech (POS) tagset and the frequency of lemma are shown below.

Table 4.2 Noun Categories

Description	Example	Frequency lemma
noun, singular or mass	grammar, communication, video, say, apology, teacher	15697
noun plural	materials, teachers, friends, students, speakers	3770
proper noun, singular	audio, video, English, morning, Mr. Mrs. Ahmad, Susan, Mia	5904
proper noun, plural	Mondays, arts, pictures, states	13

Table 4.2 above shows the frequency of nouns that have been accumulated, some categories are the following.

- noun, singular or mass, for example: grammar, communication, video, say, apology, teacher, etc. and the frequency of lemma is 15697.
- noun plural, for example: *materials, teachers, friends, students, speakers, etc.* and the frequency of lemma is 3770.
- proper noun, singular, for example: audio, video, English, morning, Mr., Mrs., Ahmad, Susan, Mia, etc. and the frequency of lemma is 5904.
- proper noun, plural, for example: *Mondays, arts, pictures, states, etc.* and the frequency of lemma is 13.

b. Verb Categories

The categories of the "verb" in the Part of Speech (POS) tagset and the frequency of lemma are shown below.

Table 4.3 Verb Categories

Description	Example	Frequency of Lemma
verb "be", base form	be	352
verb "be", past tense	was, were	431
verb "be", gerund/present participle	being	34
verb "be", past participle	been	45
verb "be", non-3rd person singular present	am, 'm, are, 're	824
verb "be", 3rd person sing. present	is, 's	1980
verb, base form	accept, add, answer, apologize	1631
verb, present, not 3rd person	accumulates, adds, asks, arranges, becomes, comes	941
verb, past participle	accepted, added, advertised, broken, based, spent	1160
verb, gerund/present participle	according, adding, answering, asking, attending	1462
verb, past tense	wrote, won, ate, became, arrived	971
verb, base form	accept, accompany, add, access agree, answer	4554
verb "have", 3rd person sing. present	has, 's	271
verb "have", sing. present, non-3rd	have, 've	395
verb "have", past participle	had	4

verb "have", gerund/present participle	having	24
verb "have", past tense	had	82
verb "have", base form	have	101

Table 4.3 above shows the frequency of verbs that have been accumulated, the most frequency vocabulary that appears above 1000 lemma in the verb category are below.

- verb "be", 3rd person sing. present, the example: *is*, 's and the frequency of lemma is 1980.
- verb, base form, the example: *accept, add, answer, apologize* and the frequency of lemma is 1631.
- verb, past participle, the example: accepted, added, advertised, broken, based, spent and the frequency of lemma is 1160.
- verb, gerund/present participle, the example: *according*, *adding*, *answering*, *asking*, *attending* and the frequency of lemma is 1462.
- verb, base form, the example: *accept, accompany, add, access agree,*answer and the frequency of lemma is 4554.

c. Adjective Categories

The categories of the adjective in the Part of Speech (POS) tagset and the frequency of lemma are shown below.

Table 4.4 Adjective Categories

Description	Example	Frequency of Lemma
adjective	rich, inspiring, interactive, various	4961
adjective comparative	younger, longer, more, taller, larger, slower, warmer, better, worse, less, bigger	137
adjective superlative	best, most, youngest newest, closest, biggest, worst	190

Table 4.4 above shows the frequency of adjectives that have been accumulated, some categories include the following.

- adjective, the example: rich, inspiring, interactive, various etc. and the frequency of lemma is 4961.
- adjective comparative, the example: younger, longer, more, taller, larger, slower, warmer, better, worse, less, bigger etc. and the frequency of lemma is 137.
- adjective superlative, the example: best, most, youngest newest, closest, biggest, *worst etc.* and the frequency of lemma is 190.

d. Adverb Categories

The categories of the adverb in the Part of Speech (POS) tagset and the frequency of lemma are shown below.

Table 4.5 Adverb Categories

Description	Example	Frequency
adverb	above, so, about soon, now, yesterday, right, carefully	196
adverb comparative	better, earlier, further, later	83
adverb superlative	best, most	36

Table 4.5 above shows the frequency of adverbs that have been accumulated, some categories are presented below.

- adverb, the example: *above, so, about soon, now, yesterday, right,*carefully and the frequency of lemma is 196
- adverb comparative, the example: *better, earlier, further, later* and the frequency of lemma is 8.
- adverb superlative, the example: *best, most* and the frequency of lemma is 36.

e. Pronoun Categories

The categories of the pronoun in the Part of Speech (POS) tagset and the frequency of lemma are shown below.

Table 4.6 Pronoun Categories

Description	Example	Frequency
possessive ending	's	433
personal pronoun	us, we, I, you, it, him, he	5872
possessive pronoun	our, your, her, my, his	1781

Table 4.6 above shows the frequency of pronouns that have been accumulated, some categories are the following:

- possessive ending, the example: 's and the frequency of lemma is 433
- personal pronoun, the example: *us, we, I, you, it, him, he* and the frequency of lemma is 5872

- possessive pronoun, the example: *our, your, her, my, his* and the frequency of lemma is 1781
- f. Preposition, conjunction, interjection and determiner categories

 The categories of the reposition, conjunction, interjection and determiner in
 the Part of Speech (POS) tagset and the frequency of lemma are shown
 below.

Table 4.7

Preposition, conjunction, interjection and determiner categories

Part of Speech	Example	Frequency of Lemma
preposition	about, above, across, after,	6059
	against, along, although,	
	a <mark>mong, arou</mark> nd	
conjunction	and, or, but, both, either,	1921
interjection	Uhhuhhuhh, Hello, Please, Oh,	376
	Yes, Hey, No, OK	
determiner	this, some, a, the, every, all,	8173
	these, that, any	

Table 4.7 above shows the frequency of preposition, conjunction, interjection and determiner categories that have been accumulated, some categories are:

- preposition, the example: *about, above, across, after, against, along, although, among, around* and the frequency of lemma is 6059
- conjunction, the example: *and*, *or*, *but*, *both*, *either*, and the frequency of lemma is 1921.

- interjection, the example: *Uhhuhhuhh*, *Hello*, *Please*, *Oh*, *Yes*, *Hey*, *No*, *OK* and the frequency of lemma is 376.
- determiner, the example: *this, some, a, the, every, all, these, that, any* and the frequency of lemma is 8173.

B. Discussion

This section describes the results of the research related to the theory that has been presented in Chapter 2. The discussion focuses on the frequency of English vocabulary used by the English Teachers forum to construct English textbooks in Indonesia, especially in the EFL context. More explanations are discussed as follows.

This research discusses the monolingual corpus because it collects only English vocabularies that the English Teacher Forum frequently used in constructing textbooks. Besides, a synchronic corpus is discussed from the vocabulary of the textbooks that are constructed at the same point in time for the 2019/2020 academic year. In linguistics, synchronic corpus (as opposed to diachronic corpus) refers to the study of language at a single point in time, generally the present, without regard for its history or development. As a result, a synchronic corpus is constructed from currently used languages and provides data from a single point in time.

In terms of token and type, the word profiler tool of AntWordProfiler calculation of the corpora shows 82,918 tokens and 6,669 types from five English textbooks. These results are in line with the study from Asnidar, which

revealed that three coursebook books on the English textbook used in ELT for non-English majors had 5,290-word types and 63,130 tokens.

Then, the findings of this study revealed that the most commonly used words in books were nouns, verbs, and determiners (see figure 4.1). This shows that the words in those categories were considered as a priority for student learning. Furthermore, the data reveals that many words from textbooks are used in ordinary speech, such as audio, video, English, morning, Mr. Mrs. Ahmad, Susan, and Mia in noun categories (see table 4.2). It also reveals that the number and types of words used in the texts differ significantly. A few examples include a, the, information, sentences, thank, birthday, and tomorrow. It suggests that word selection in textbooks does not appear to be based on common thinking that the English teacher forum in Tuban has constructed. Therefore, knowing the information about the most vocabulary of the textbook was important, and it was helpful for the teachers to check the vocabulary in the material that would be given to the students. Furthermore, this was good to help the teacher in choosing suitable vocabulary assignments, exercises, homework, and many other activities by knowing the difficulties of the vocabulary in each chapter of the textbooks.

In addition, High-Frequency Words (HPW) are words that appear frequently in a variety of contexts. The majority of the terms in HFW are content and function words. Nouns, verbs, adjectives, and adverbs are examples of content words. Then, pronouns, determiners, and prepositions are examples of function words. Because the results in this study showed that most frequency

words, namely nouns, verbs, determiners, the words used in the textbooks developed by the teachers are within the general category of HFW. The words that occur very frequently in this study are also included in the High-Frequency Words. These words were very important, Nation also recommended that these words be included on the first list of the target words in any type of vocabulary development program.⁵⁷ These findings indicate that the corpora in the books use the important words that should be prioritized in the target words to facilitate students' vocabulary development in the Indonesian EFL context, particularly those in the Tuban district.

Moreover, according to Nation, Michael West's General Service Lists (GSL) featured roughly 2000 words families as the most popular High-Frequency Words (HFW). In this research, the HFW of the textbooks that have been analyzed contains the same vocabularies in GSL (see appendix 1). Only a few words were not in the GSL but appeared in the High-Frequency Words of the textbooks. It was classified as a number (eight, nine, six, twenty, three, four, five, thirteen, twelve, thirty, eleven), name of the day (Monday, Thursday, Wednesday, Sunday), name of the month (February, August, September, October, November).

As a result, these words should be prioritized because it is impossible to understand English without knowing these words. High-frequency words or general service words are worth paying attention to because they frequently

7 -

⁵⁷ Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

appear in readings. Teachers can deal with these words by pre-teaching them, incorporating them into a text-based exercise, or spending time looking at the different meanings they have. This set of words is clearly very useful and important for English learners.

Regarding the AWL word families, the AWL coverage of the textbooks contained 282-word families and accounted for about 2.51% of the tokens among the 570 AWL word families (see table 4.1). The result did not meet Coxhead's 10% coverage of academic texts. On the one hand, the text coverage of the AWL word families in English majors' university textbooks corpus was 6.5%. On the other hand, it was also much lower than the 11.7% of the AWL words in a study done by Vongpumivitch and et al. that tried to explore the frequency of Coxhead's AWL word families in their 1.5 million words corpus of Applied Linguistics Research Articles.

Furthermore, when these findings were compared to earlier research, it was shown that coverage of the AWL items in five English textbooks is lower than in other fields. This study has a lesser AWL coverage than Mehri Jamalzadeh and Azize Chalak Physiotherapy Research Articles. According to the analysis, 562 of the 570-word families in AWL were often found in physiotherapy research articles (RAs), accounting for 11.51 percent of the tokens in the corpus.

This finding is also considerably lower than the AWL coverage of 13.12% of the tokens in the corpus reported by Ismail & Nasrin Xodabande for psychology research articles. These findings indicate that when the teacher

forum constructed the five English textbooks, the academic vocabularies were not their priority. This is in line with what Coxhead et al. said about ELT, GE, and middle school texts. However, they include little academic terminology because they are used for other reasons in language acquisition or because they are written in a "relatively reader-friendly" manner and thus are not as scholarly as tertiary texts or research papers.

Junior high school students are not emphasized using academic words when they write a text in the teaching and learning process. So, academic words in this research are not entirely useful for the student in junior high school because the words in the AWL category are mostly used only in academic publications. In addition, the students at least get new knowledge of academic words because they are used in textbooks. This research is in line with the research about the Academic Word List (AWL) conducted by Mehri Jamalzadeh and Azize Chalak. The words are not entirely helpful for the physiotherapy learners because some word families have narrow coverage besides the lack of using academic words frequently.

As pedagogical implications, the examination of corpus data on vocabulary frequency that teachers used for constructed textbooks provides valuable insights. The current research provides teachers with an understanding of which vocabulary they need to teach students to meet the curriculum expectations. In a range of spoken and written texts, knowing high-frequency and scholarly words is vital. It can aid students in their recognition of a huge number of words. Knowing the academic word list is particularly significant

since it helps pupils improve their reading comprehension and serious tone of writing. In other words, students will find it easier to grasp reading and writing comprehension in academic texts if they are familiar with high-frequency and academic vocabulary.

The advantages of adding vocabulary to pupils will make teaching English easier for teachers. As a result, the learning process will be more effective because it will concentrate on important words. Then, the teacher can develop a three-year teaching plan for junior high school students that can make vocabulary learning more controlled and thus more successful.

CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

The corpus study examines the English vocabulary that is frequently used by the Junior High School English teacher forum when constructing the textbook in Tuban. Five textbooks constructed by the English Teacher Forum in Tuban have been analyzed and yielded important results. The corpus analysis of selected core textbooks contains 82,918 tokens and 6,669 types. Then, the most frequent words used included High Frequently Word, namely nouns, verbs, and determiners. It should be prioritized in the target words to facilitate students' vocabulary development in the Indonesian EFL context, particularly those in the Tuban district. Besides, the textbooks contained 282-word families of academic words and accounted for about 2.51% of the tokens. The current research provides teachers with an understanding of which vocabulary they need to teach students to meet the curriculum expectations. Then, English language teaching programs will become more efficient.

B. Suggestion

1. For the teacher

English teachers are encouraged to be familiar with the most frequently used vocabularies in their students' textbooks in terms of English vocabulary teaching. Maybe the teachers have the list of vocabulary taken from the textbook that contains words that are indicated as difficult words.

Hence, the teachers can provide alternative words or synonyms of those difficult words. By having that kind of thing, it was good to help the teachers give a suitable explanation in the classroom that may help the students easily understand the main point of the materials.

2. For ELT material developers

It is recommended that ELT materials developers use empirical corpusbased data when developing materials, particularly when selecting texts and grading vocabulary. They have the option of supplementing this corpus-based information with their subjective judgments via a mixedmethod explanatory design or using it exclusively to develop ELT vocabulary materials. As a result, ELT in its specific context will be constantly updated to reflect current linguistics and ELT research developments.

3. For the further researcher

Combining quantitative and qualitative measures to generate a frequency word list remains an intriguing research inquiry, as it challenges the two distinct poles of excessive reliance on quantitative or qualitative criteria. The use of word lists in informed materials development is also a relatively unexplored area. As a result, it is recommended that additional research on corpus-based research be conducted with the goal of pedagogical benefits, particularly for ELT in Indonesia.

REFERENCES

- Anne O'Keeffe, Michael Mccarthy, and Ronald Carter, "From Corpus to Classroom: Language Use and Language Teaching" (n.d.): 1.
- Bennett, Gena R. 2010. Using Corpora in the Language Learning Classroom. Michigan ELT.
- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? Studies in Second Language Acquisition, 24(2), 199–208.
- Budi Setiawan, 'Improving The Students Vocabulary Mastery through Direct Instruction' (Universitas Negeri Surakarta., 2010).
- Burton, G. (2012). Corpora and coursebooks: destined to be strangers forever? Corpora, 7(1), 91–108.
- Christine Dugan, Strategies for Building Vocabulary in Language Arts (Huntington beach: Shell Education publishing Inc, 2010).
- Content and function words in sentence stress, Retrieved from: https://pronuncian.com/content-and-function-words
- Dharma, Y. P., & Aristo, T. J. V. (2018). An analysis of English textbook relevance to the 2013 English curriculum. Journal of English Educational Study (JEES), 1(1), 24-33.
- Foster, J., & Mackie, C. (2013). Lexical analysis of the Dr. Seuss Corpus. Concordia Working Papers in Applied Linguistics, 4, 1–21.
- Garinger, D. (2001). Textbook evaluation. TEFL Web Journal.
- Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), Computer learner corpora, second language acquisition and foreign language teaching (pp. 3–33). Philadelphia: John Benjamins.
- Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), Computer learner corpora, second language acquisition and foreign language teaching (pp. 3–33). Philadelphia: John Benjamins.
- Graves, K. (2003). Coursebooks. In Nunan, D. Practical English language teaching. New York: McGraw Hill Inc.
- Grigaliuniene, Jone. 2013. Corpora in Language Studies. Lituania: Vilnius University.
- Grigaliuniene, Jone. 2013. Corpora in Language Studies. Lituania: Vilnius University.
- Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. Ampersand, 2, 136–144.
- Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a Corpus-Based Word List in Pharmacy Research Articles: A Focus on Academic Culture. International Journal of Society, Culture & Language, 8(1), 1-15.
- Hoey, M. (2000). The hidden lexical clues of textual organization: A preliminary investigation into an unusual text from a corpus perspective. In L. Burnard & T. McEnery (Eds.), Rethinking language pedagogy from a corpus perspective (pp. 31–42). New York: Peter Lang.

- Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage, Electronic Journal of Foreign Language Teaching, 6(1), 42–62.
- https://www.sketchengine.eu/corpora-and-languages/corpus-types/
- Hunston, S., & Francis, G. (1999). Pattern grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.
- Jamalzadeh, M., & Chalak, A. (2019). A Corpus-based study of academic vocabulary in physiotherapy research articles. Language Teaching Research Quarterly, 9(1), 69-82.
- John W, Cresswel. Qualitative Inquiry And Study Design: Choosing Among Five Approaches Second Edition (New Delhi. 2007),156.
- Kaewphannagam, C., Broughton, M.M., & Soranasataporn, S. 2002. Corpus-Based Analysis: Guidelines for Getting Practical Language Input in Materials Development.
- Keck, C.M. (2004). Corpus linguistics and language teaching research: Bridging the gap. Language Teaching Research, 8(1), 83–109.
- Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. The Language Learning Journal, 40(1), 35–45.
- Lindquist, H. 2009. Corpus Linguistics and the Description of English. Edinburgh: Edinburgh University Press.
- Lotfi, Ghazal, Learning Vocabulary in EFL Context Through Vocabulary Learning Service, Novitas Royal, Vol. 1, no. 2, 2007.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. Language Teaching, 41(4), 563–574.
- McKenery, T. & Wilson, A. (2001) Corpus Linguistics. Edinburgh: Edinburgh UK, Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge:
- Cambridge University Press.
- Oxford Advanced Learner's Dictionary (9th edition): Oxford University Press
- Pathan, H., Memon, R. A., Memon, S., Shah, S. W. A., & Magsi, A. (2018). Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups. International Journal of English Linguistics, 8(4).
- Richards, J. C. (2001). Curriculum development in language teaching. Ernst Klett Sprachen.
- Sardinha, T.B. (2004). Linguística de corpus. Barueri, SP: Manole.
- Shin, D. & Chon, Y.V. (2011). A corpus-based analysis of curriculum-based elementary and secondary English textbooks. Multimedia-Assisted Language Learning, 14(1), 149-175.
- Skolverket. (2006). The role of teaching materials in teaching: Secondary school teachers' choice, use and evaluation of teaching materials in Arts, English and Social Sciences. Stockholm: Skolverket.
- Sudarman, C. S. 2018. The English Vocabulary Size and Level of English Department Students at Kutai Kartanegara University. Journal UIN Alauddin, 4(1)

- Szudarski, P. (2011). The Role of Language Corpora in Teaching English as a Foreign Language in Poland.
- Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), Rethinking language pedagogy from a corpus perspective (pp. 91–102). New York, NY: Peter Lang.
- Ur, P. (1996). A course in language teaching : practice and theory. Cambridge : Cambridge University Press.
- Ur, P. (1996). A course in language teaching : practice and theory. Cambridge : Cambridge University Press.
- Wang, P. (2017). A corpus-based study of English vocabulary in art research articles. Journal of Arts and Humanities, 6(8), 47-53.
- West, M. (1953). A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology. London: Longman, Green & Co.
- What Is a Corpus, What Is Corpus Linguistics?" (n.d.): 3.
- Zahrotul Izzah, 'The Use of Word Chain Game to Improve Vocabulary Mastery Of Eight Grade Students at SMPN 3 Kalibagor' (Universitas Muhammadiyah Purwokerto, 2015).