

**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



**UIN SUNAN AMPEL
S U R A B A Y A**

By:

Siti Uswatun Khasanah

NIM: D05217021

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

2021

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 pernyataan tidak sesuai dengan fakta yang ada, maka saya selaku penulis bersedia dimintai pertanggung jawaban.

Surabaya, 3 Juli 2021

Pembuat pernyataan



Siti Uswatun Khasanah

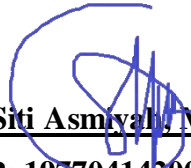
D05217021

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 Asmyah, M.TESOL

NIP: 197704142006042003

Advisor II,



Rizka Satriyani, 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,

Prof. Dr. H. Ali Mas'ud, M. Ag., M. Pd

NIP. 196301231993031002

Examiner I,

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

NIP. 197310131997031002

Examiner II,

Dr. Irma Soraya, M.P.d

NIP: 196709301993032004

Examiner III,

Dr. Siti Asmiyah, M. TESOL

NIP: 197704142006042003

Examiner IV,

Rizka Satriyani, 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 akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Siti Uswatun Khasanah
NIM : D05217021
Fakultas/Jurusan : Fakultas Tarbiyah dan Keguruan / Pendidikan Bahasa Inggris
E-mail address : sitiuswatunkha@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)
yang berjudul :

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

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 8 Juli 2021

Penulis

(Siti Uswatun Khasanah)

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

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

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.

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

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.

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

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

- 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)

