

**CORPUS ANALYSIS OF ACADEMIC VOCABULARY IN
STUDENTS' THESIS PROPOSAL AT ENGLISH TEACHER
EDUCATION DEPARTMENT OF UIN SUNAN AMPEL
SURABAYA**

THESIS

**Submitted in partial fulfilment of the requirement for the degree of
Sarjana Pendidikan (S.Pd) in Teaching English**



**By:
M. Sahilur Rozaq Mahmudi
NIM: D75213080**

**ENGLISH LANGUAGE EDUCATION DEPARTMENT
FACULTY OF TARBIYAH AND TEACHER TRAINING
UIN SUNAN AMPEL SURABAYA
SURABAYA
2019**

PERNYATAAN KEASLIAN TULISAN

Yang bertanda tangan dibawah ini

Nama : M. Sahilur Rozaq Mahmudi
NIM : D75213080
Semester : XII
Fakultas/Prodi : Tarbiyah dan Keguruan/Pendidikan

Bahasa Inggris

Dengan ini menyatakan sebenar-benarnya bahwa skripsi yang berjudul *“Corpus Analysis of Academic Vocabulary in Students’ Thesis Proposal at English Teacher Education Department of UIN Surabaya”* adalah benar-benar merupakan hasil karya sendiri. Segala materi yang diambil dari karya orang lain hanya digunakan sebagai acuan dengan mengikuti tata cara dan etika penulisan karya ilmiah yang ditetapkan oleh jurusan.

Demikian pernyataan ini dibuat dengan sebenar-benarnya apabila pernyataan tidak sesuai dengan fakta yang ada, maka saya selaku penulis bersedia dimintai pertanggungjawaban sesuai ketentuan peraturan perundang-undangan yang berlaku.

Surabaya, 03 Agustus 2019

Pembuat pernyataan



Rozq Mahmudi

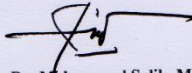
NIM: D75213080

ADVISOR APPROVAL SHEET

This thesis by M. Sahilur Rozaq Mahmudi entitled "*Corpus Analysis of Academic Vocabulary in Students' Thesis Proposal at English Teacher Education Department of UIN Sunan Ampel Surabaya*" has been approved by the thesis advisor for further approval by the Boards of Examiners.

Surabaya, July 23rd 2019

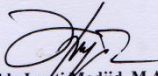
•
Advisor I,



Dr. Mohammad Salik, M.Ag.

NIP. 196712121994031002

Advisor II,



Hilda Izzati Madjid, M.A

NIP. 198602102011012012

EXAMINER APPROVAL SHEET

This thesis by M. Sahilur Rozaq Mahmudi entitled "*Corpus Analysis of Academic Vocabulary in Students' Thesis Proposal at English Teacher Education Department of UIN Surabaya*" has been examined on July 16th 2019 and approved by the Board of Examiners.



Dean,

[Signature]

M. Mas'ud, M.Ag. M.Pd.I
NIP. 296301231993031002

The Board of Examiners
Examiner I,

[Signature]
Rizka Saffriyani, M.Pd
NIP. 198409142009122005

Examiner II,

[Signature]
Drs. Muhtarom, M.Ed. Grad. Dipl. TESOL
NIP. 196512201992031005

Examiner III,

[Signature]
Hilda Izzati Madjid, M.A
NIP. 198602202011012012

Examiner IV,

[Signature]
Rakhmawati, M.Pd
NIP. 197803172009122002



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpustakaan@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 : MOHAMMAD SAHILUR ROZAQ MAHMUDI
NIM : D75213080
Fakultas/Jurusan : TARBIYAH / PENDIDIKAN BAHASA INGGRIS
E-mail address : LRI.RYUZAKI@GMAIL.COM

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

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

CORPUS ANALYSIS OF ACADEMIC VOCABULARY IN STUDENTS' THESIS
PROPOSAL AT ENGLISH TEACHER EDUCATION DEPARTMENT OF UIN SUNAN
AMPEL SURABAYA

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,

Penulis


(M. Sahilur Rozaq)
nama terang dan tanda tangan

ABSTRACT

Rozaq, Mohammad Sahilur. 2019. *Corpus Analysis of Academic Vocabulary in Students' Thesis Proposal Seminar Class at English Teacher Education Department of UIN Sunan Ampel Surabaya*. A thesis. English Teacher Education Department, Faculty of Education and Teacher Training, Sunan Ampel State University. Surabaya. Advisors: Dr. Muhammad Salik M.Ag and Hilda Izzati Madjid, M.A

Key Word: *Academic Vocabulary, Thesis Proposal, Corpus Analysis, Word Frequency*

This thesis discusses the academic vocabulary used in thesis proposal by seventh semester students. The problem of this research is how academic vocabulary used by seventh semester students in their thesis proposal. This research is conducted using quantitative method. In order to achieve the goal, the researcher comprises 15 thesis proposal documents from seventh semester students in academic year 2018/2019 and the thesis proposal were analysed through an application called AntWordProfiler by Anthony Laurence. The researcher uses Academic Vocabulary List (AWL) compiled by Averil Coxhead as an instrument and as a standardized tool of measurement that included to the application. The result of this research based on the finding and discussion; this research figured out that academic vocabulary used by seventh semester students are quite high with 11.92% academic word list token from 66419 of running words. Specifically, there are fourteen out of fifteen analysed documents are including to academic work / journal. Since, the academic word list has to cover approximately 10% in a piece of academic document. Furthermore, students' knowledge about academic vocabulary can still continue to grow as long as they learn and practice using academic vocabulary in their work.

ABSTRAK

Rozaq, Mohammad Sahilur. 2019. *Corpus Analysis of Academic Vocabulary in Students' Thesis Proposal Seminar Class at English Teacher Education Department of UIN Sunan Ampel Surabaya*. A thesis. English Teacher Education Department, Faculty of Education and Teacher Training, Sunan Ampel State University. Surabaya. Advisors: Dr. Muhammad Salik M.Ag and Hilda Izzati Madjid, M.A

Key Word: *Academic Vocabulary, Thesis Proposal, Corpus Analysis, Word Frequency*

Skripsi ini mendiskusikan tentang penggunaan kosakata akademik pada proposal skripsi mahasiswa semester tujuh. Penelitian ini membahas tentang kosakata akademik yang digunakan pada proposal skripsi oleh mahasiswa semester tujuh. Penelitian ini membahas tentang permasalahan bagaimana penggunaan kosakata akademik oleh semester tujuh pada proposal skripsi mereka. Penelitian ini menggunakan metode quantitative. Untuk mencapai tujuan dari penelitian ini, peneliti membandingkan 15 dokumen proposal skripsi dari mahasiswa semester tujuh pada tahun ajaran 2018/2019 dan proposal skripsi ini dianalisa menggunakan aplikasi yang dinamakan AntWordProfiler oleh Anthony Laurence. Peneliti juga menggunakan daftar kosaka akademik yang diadopsi dari Averil Coxhead sebagai instrument dan sebagai standar pengukuran dimana daftar kosakata tersebut terdaftar dalam aplikasi AntWordProfiler. Hasil dari penelitian ini menyimpulkan bahwa kosakata akademik yang digunakan oleh mahasiswa semester tujuh cukup tinggi dengan presentase kosakata akademik 11.92% dari 66419 jumlah kata keseluruhan. Lebih jelasnya ada 14 dari 15 dokumen yang dianalisa termasuk kedalam kategori jurnal akademik, dimana kosakata akademik harus mencakup paling tidak 10% dari sebuah dokumen atau jurnal akademik. Selebihnya, pemahaman mahasiswa tentang kosakata akademik harus terus bertambah selama mereka belajar dan berlatih dalam menggunakan kosakata akademik dalam pekerjaan mereka.

TABLE OF CONTENT

TITTLE	i
PERNYATAAN KEASLIAN TULISAN.....	ii
ADVISOR APPROVAL SHEET	iii
EXAMINER APPROVAL SHEET	iv
LEMBAR PERSETUJUAN PUBLIKASI.....	v
ABSTRACT	vi
ABSTRAK	vii
TABLE OF CONTENT	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF APPENDICES	xiii
LIST OF ABREVIATION	xiv
CHAPTER I.....	1
INTRODUCTION	1
A. Background of the Study.....	1
B. Statement of Research Problem.....	5
C. Objective of the Study.....	6
D. Significance of the Study	6
1. For the Students	6
2. For the Researcher	6
3. For the Reader.....	6
E. Scope and Limitation of the Study.....	6
1. The Scope of the Study	6
2. Limit of the Study	7
F. Definition of Key Term.....	8

1. Corpus	8
2. Academic Vocabulary	8
3. Thesis Proposal	9
CHAPTER II	10
REVIEW OF RELATED LITERATURE	10
A. Theoretical Foundation	10
1. Corpus Analysis	10
a. Definition of Corpus Analysis.....	10
b. Types of Corpora	17
2. Writing	19
a. Definition of Writing	19
b. Academic Writing.....	19
3. Vocabulary	25
a. Definition of Vocabulary	25
b. Types of Vocabulary.....	26
4. Vocabulary Usage.....	27
5. Academic Vocabulary	28
a. Definition of Academic Vocabulary.....	28
B. Previous Study	33
CHAPTER III.....	36
RESEARCH METHODOLOGY	36
A. Approach and Research Design	36
B. Subject of the Study and Setting.....	36
1. Population.....	36
2. Sample.....	36
C. Setting of the Study.....	37

D.	Data and Source of Data.....	37
E.	Data Collection Technique.....	37
F.	Research Instrument.....	38
G.	Data Analysis Technique.....	38
H.	Research Stages.....	39
CHAPTER IV.....		41
RESEARCH FINDINGS AND DISCUSSION.....		41
A.	Research Findings.....	41
1.	Academic Vocabulary Used in Proposal Writing.....	41
a.	Analysis.....	45
b.	Data.....	45
c.	Method.....	45
d.	Definition.....	45
e.	Process.....	46
f.	Research.....	46
g.	Researcher.....	46
h.	Scope.....	46
i.	Source.....	46
j.	Technique.....	47
2.	Data Analysis on Students' Thesis Proposal.....	48
B.	Discussion.....	50
1.	Academic Vocabulary Used in Proposal Writing.....	50
2.	Data Analysis on Students' Thesis Proposal.....	50
CHAPTER V.....		53
CONCLUSION AND SUGGESTION.....		53
A.	Conclusion.....	53

B. Suggestion	53
REFERENCES	56

LIST OF TABLES

No.	NAME	PAGE
1.	Table 3.1: The Example of Result of AntWordProfiler	39
2.	Table 4.1: The Result of Calculation	42
3.	Table 4.2: Vocabulary Used by all Participants	43
4.	Table 4.3: Percentage Token of Academic Vocabulary	47

LIST OF FIGURES

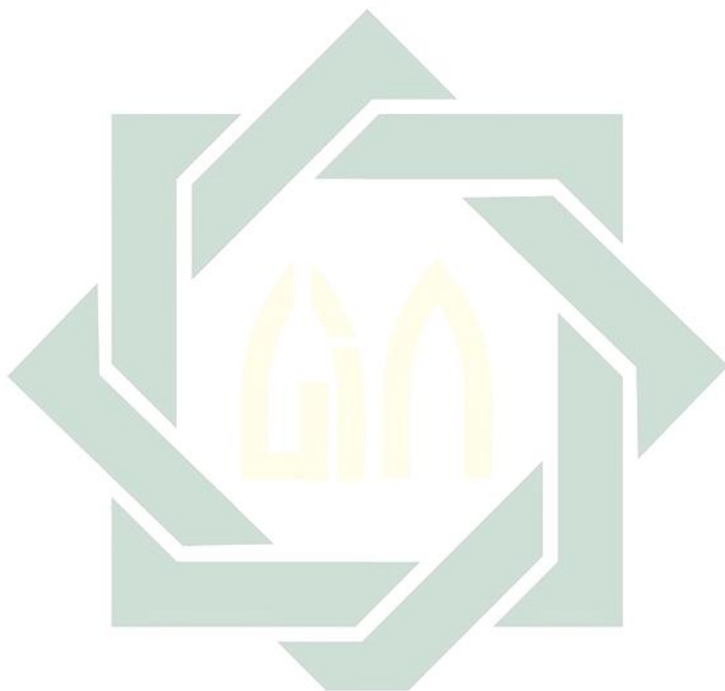
No.	NAME	PAGE
1.	Figure 2.1: AntConc main window	13
2.	Figure 2.2: WordSmith main window	14
3.	Figure 2.3: AntWordProfiler main window	15
4.	Figure 2.4: Part of Essay	21
5.	Figure 2.5: Types of Vocabulary	30
6.	Figure 4.1: Academic Word List Rank	50

LIST OF APPENDICES

Appendix 1 Percentage of Academic Word List

Appendix 2 Spreading Data of Academic Vocabulary

Appendix 3 Sub List of Academic Word List



LIST OF ABBREVIATION

L1	= First Language
L2	= Second Language
EFL	= English as Foreign Language
ESL	= English as a Second Language
GSL	= General Service List
UIN	= Universitas Islam Negeri
KWIC	= Keyword-In-Context
CIC	= Cambridge International Corpus
COBUILD	= Collins Birmingham University international language database
COLT	= The Bergen Corpus of London Teenager Language
BNC	= British national corpus
CANCODE	= Cambridge and Nottingham Corpus of Discourse in English
LOCNEE	= Louvain Corpus of Native English Essays
ICLE	= International Corpus of Learner English
AWL	= Academic Word List
CCSS	= Common Core State Standard
IEEC	= Information Engineering English Corpus
IJAL	= Indonesian Journal of Applied Linguistic
OUP	= Oxford University Press
PAR 1 – 15	= Participant 1 - 15

CHAPTER I

INTRODUCTION

This chapter discusses the introduction of the study that explain the reason of conducting this research and research questions that would bring this research to some eases. Objectives of the study that present the aims of this research and significances of the study are also provided in this chapter. Furthermore, this chapter presents scope and limitation and definition of key terms that can explain each variable that is used in this research.

A. Background of the Study

Vocabulary is one of the important components of language acquisition because with vocabulary people can understand and produce the second language learned. People also use language to communicate world widely, so naturally the problem in vocabulary studies is how much vocabulary is necessary to enable this communication. The short answer is a lot, but it depends on their learning goals. If one wishes to achieve native-like ability, then presumably it is necessary to have a vocabulary size similar to native speaker.¹ The more vocabulary acquired by people, the easier they will master the language. Vocabulary acquisition is incremental both in terms of acquiring and adequate vocabulary size.² A research has done by Henriksen, she measured the second language (L2) vocabulary size of Danish EFL students, she also measured their first language (L1) size. The result found that there was consistent improvement in vocabulary size across the increasing grade in both L1 and L2. Unsurprisingly, she also found that the L1 scores were larger than the L2 scores, even though the L1 test was very low frequency items compared to the L2 test.³

¹ Norbert Schmitt, *Researching Vocabulary : A Vocabulary Research Manual*, Research and practice in applied linguistics (Houndmills, Basingstoke, Hampshire : Palgrave Macmillan, 2010), 6.

² *Ibid.*, 19.

³ Dorte Albrechtsen, Kirsten Haastrup, and Birgit Henriksen, *Vocabulary and Writing in a First and Second Language* (London: Palgrave Macmillan UK, 2008), 57, accessed August 20, 2019, <http://link.springer.com/10.1057/9780230593404>.

The vocabulary can be acquired in some ways such as reading and listening. But acquiring vocabulary from listening and reading are not always in academic vocabulary form. For example, some vocabulary that acquire from song, comic and movie, their vocabulary form usually using slang, common language or daily language.

In other hand, students nowadays rarely think about what kinds of vocabulary they used in their written or spoken assignments, they just pour the ideas to finish the assignments. Whereas, vocabulary used in university frequently using academic form because there are some assignments that tend to use academic vocabulary for example write a journal, speech text and write a thesis which they are consumed by public as a reference. Because academic vocabulary has a specific function rather than daily or common vocabulary to express academic communication. For this reason, the mastery of academic vocabulary is beneficial and urgent for university students.⁴

However, a research has found that most of the students in university especially non-native speaker of English university have some difficulties in using the academic vocabulary because of the insufficient vocabulary knowledge, such as what words or types of words occur in this sentence, or where, when and how often can we use this word. For example Evans and Green⁵ they have been surveyed Cantonese speaking students at Hong Kong's largest English-medium university and the result concluded that "inadequate receptive and productive vocabulary in English is the main problem confronting the almost 5000 students who participated in the survey".⁶ This condition is similar with Indonesian students where English is not their native language, when writing a thesis or journal they will face some difficulties in using academic vocabulary.

⁴ Hans Malmström, Diane Pecorari, and Philip Shaw, "Words for What? Contrasting University Students' Receptive and Productive Academic Vocabulary Needs," *English for Specific Purposes* 50 (April 2018): 2.

⁵ Stephen Evans and Christopher Green, "Why EAP Is Necessary: A Survey of Hong Kong Tertiary Students," *Journal of English for Academic Purposes* 6, no. 1 (January 1, 2007): 3-17.

⁶ *Ibid.*

Ideally writing a thesis or journal have to use academic vocabulary, but in fact students of language teaching in university still unfrequently to apply it in their academic assignments, for example most of the students using word *get* instead of *acquiring* in academic context, however, they are supposed to be able to differentiate when to use academic language and common language. Another reason why they do not use academic language is that they have limited knowledge about academic language.

There has been some research about academic vocabulary. As an example, Eniko Csomay conduct a study about academic vocabulary in ESL student paper. The result found that significant relationship between academic vocabulary use and essay scores in some text-types.⁷ Furthermore, Hans Malmstrom conducted the research about contrasting university students' receptive and productive academic vocabulary needs. The result showed that productive and receptive academic vocabulary needs are far from identical which mean the student rarely use academic vocabulary in their both writing and speaking.⁸ Similarly, Elisabeth Schuth conducted the study about the influence of academic vocabulary knowledge on school performance and the result support the claim that academic language proficiency significantly influences academic carrier.⁹ Another research is about vocabulary usage by Lili Zhai. Her research was conducted on Chinese EFL Learners. She focuses on learners' vocabulary proficiency, especially vocabulary in English writing. Finished one composition for analysing subject's vocabulary usage, i.e. lexical richness includes lexical sophistication and lexical variation.¹⁰ Furthermore, Mehri Jamalzadeh and Azize Chalak conduct a research about corpus-based study of academic vocabulary in Physiotherapy. In this study, they compared a list of academic words used frequently in physiotherapy research articles

⁷ Eniko Csomay and Alexandra Prades, "Academic Vocabulary in ESL Student Papers: A Corpus-Based Study," *Journal of English for Academic Purposes* 33 (May 2018): 1.

⁸ Malmström, Pecorari, and Shaw, "Words for What?," 1.

⁹ Elisabeth Schuth, Judith Köhne, and Sabine Weinert, "The Influence of Academic Vocabulary Knowledge on School Performance," *Learning and Instruction* 49 (June 2017): 1.

¹⁰ Lili Zhai, "A Study on Chinese EFL Learners' Vocabulary Usage in Writing," *Journal of Language Teaching and Research* 7, no. 4 (July 1, 2016): 752.

with the distribution of high frequency words in Coxhead's AWL and West's General Service List (GSL). The result indicated that AWL is not entirely useful for physiotherapy learners because of the narrow coverage of some word families and the shortage of frequently used Physiotherapy academic words.¹¹ Another research related to academic vocabulary has been conducted by Le Pham Hoai Huong from Hue University of Foreign Languages in Vietnam. The study was set out to investigate EFL university students' strategies for learning academic English words. The findings reveal that the respondents tended to use on-line dictionaries and other application more than cognitive strategies in learning academic words.¹² To conclude the research mention above focuses on how the academic vocabulary knowledge influence the academic performance.

In this research, the researcher focuses on corpus analysis as one of the most significant developments in vocabulary studies. In recent years corpus analysis has been used as evidence to provide an experimental basis for determining vocabulary behaviour, instead of relying on appeals to intuition or tradition.¹³ Several scholars (Harold Palmer, Michael West, Edward Thorndike, Lawrence Faucet, Irving Lorge) were concerned to systematize the selection of vocabulary for learners. The list ends up having about 2000 words and was finally published as the *General Service List of English Words* (GSL).¹⁴ A main feature of the GSL is that each word has different part of speech and different meaning senses which makes the list much more useful in frequency count.

UIN Sunan Ampel Surabaya is one of the state university in Surabaya that has an English Teacher Education Department. The students of English Teacher Education Department of UIN Sunan Ampel should be able to use academic writing since they take the thesis proposal seminar class and in the end of the class, they have to submit the proposal for the final exam. Furthermore, according to

¹¹ Mehri Jamalzadeh and Azize Chalok, "A Corpus-Based Study of Academic Vocabulary in Physiotherapy Research Articles" (n.d.): 15.

¹² Le Pham Hoai Huong, "A Survey Study on Academic Vocabulary Learning Strategies by EFL University Students," *Journal of Language Teaching and Research* 9, no. 5 (September 1, 2018): 1009.

¹³ Schmitt, *Researching Vocabulary : A Vocabulary Research Manual*, 12.

¹⁴ Michael West, "General Service List of English," n.d.

English Teacher Education Department of UIN Sunan Ampel curriculum, students in this department have to pass some required subjects. There are several subjects that students must take in sequent. It means that another required subject has to be done before take the next higher subject. As one of the required subjects, writing proposal class can be obtained after completing other important subject such as paragraph writing, essay writing, argumentative writing and academic writing. On the other hand, the researcher never has a subject matter in terms of learning academic vocabulary used. From the reason above, the researcher is interested in analysing the proposal from this semester. This population was chosen based on the fact that they have been suggested to use the academic vocabulary since the beginning of semester when they are in paragraph writing course, because it can be a preparation to write the thesis as the final academic task for undergraduate.

In short, this research would identify students' academic vocabulary used in their thesis proposal in English Teacher Education Department. The difference with the previous studies is that this research focuses on investigating academic vocabulary used using Academic Word List¹⁵ as a standardizing measurement, AntWordProfiler and AntConc¹⁶ application as the measurement tool. Furthermore, the researcher conducted this research is to make students more consider in using academic vocabulary to develop their writing ability in an academic work or journal.

B. Statement of Research Problem

Based on the background of the study above, this study answers the following question:

How is the academic vocabulary use of seventh semester students in their thesis proposal?

¹⁵ Averil Coxhead, "An Introduction to the Academic Word List" (n.d.): 9.

¹⁶ Anthony Laurence, "Laurence Anthony's AntConc," accessed August 20, 2019, <https://www.laurenceanthony.net/software/antconc/>.

C. Objective of the Study

According to the statement of the problem above, the purpose of this study is to investigate the academic vocabulary of seventh semester students used in their thesis proposal writing.

D. Significance of the Study

The findings of this study can be significantly beneficial in these following aspects:

1. For the Students

The findings of this study will help students of English Teacher Education Department and especially for seventh semester students who enrol in thesis proposal seminar class to recognize how broad the students are using the academic vocabulary in their thesis proposal.

2. For the Researcher

The result of this study also gives many impacts for the researcher. As an English teacher in the future, the researcher can inform about the importance of using academic vocabulary in journals, articles or books which is consumed by public as a reference.

3. For the Reader

The result of this study may be able to be as reference for the reader to get information about academic vocabulary, how to use an application that calculate the amount of academic vocabulary, and as a reference also.

E. Scope and Limitation of the Study

This study focused and limited in these following aspects:

1. The Scope of the Study

The scope of this study focuses on seventh-semester students in academic year 2018/2019 of English teacher Education Department in UIN Sunan Ampel who enrolled in thesis proposal seminar class. Total students of thesis proposal

seminar class are 113 and divided into 5 classes; A, B, C, E, F and researcher uses simple random sampling to select a sample of individuals who are representative of the entire population. So, the researcher takes 3 participants in each class, so total participants are 15 which means the researcher analyses 15 thesis proposal documents.

2. Limit of the Study

This research focuses solely on academic vocabulary used in the thesis proposal document written by seventh semester students in academic year 2018/2019 who enrolled on thesis proposal seminar course. It means, the researcher crosscheck how is the academic vocabulary used according to on what word class that academic vocabulary by seventh semester in their proposal document. Some participants may put the vocabulary in appropriate structure but some participants also may use the vocabulary inappropriately. The researcher only analyzes the context of the vocabularies which specifically have 15 range based on their word class, because with 15 range it means that the vocabulary is used by all of the participants.

The researcher did not analyze the vocabulary in semantically or in form of meaning. This research identifies the total number of academic vocabularies used by seventh semester students of English Teacher Education Department of UIN Sunan Ampel Surabaya, it is not about collocation of the word because collocation is that behaviour of the language by which two or more words go together to form a unity of meaning¹⁷, but in this research it is only about the vocabulary with the next word and did not form a unity of meaning. Researcher calculates the academic vocabulary using AntWordProfiler application by Laurence Anthony¹⁸ with Academic Word List as the instrument of academic vocabulary.

¹⁷ “Collocation - Wikipedia,” accessed August 12, 2019, <https://en.wikipedia.org/wiki/Collocation>.

¹⁸ Laurence Anthony, “Laurence Anthony’s AntWordProfiler,” accessed June 1, 2019, Anthony, L. (2013). AntWordProfiler (Version 1.4.0) Windows. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.

F. Definition of Key Term

To avoid misunderstanding or misinterpretation, the key terms used in this study are defined below.

1. Corpus

Corpus (or Corpora: plural) is a collection of linguistic data, compiled as written texts or as a transcription of recorded speech, and usually stored as an electronic database. Furthermore, computer corpora may store many millions of running words, whose features can be analysed by means of tagging and the use of concordancing programs.¹⁹ In this research, corpus means a collection of written vocabularies that can be used as a standardizing of measurement of academic vocabulary.

2. Academic Vocabulary

A lot of expert provides an explanation about Academic vocabulary. Firstly, Barber argue that Academic Vocabulary is variously known as generally useful scientific vocabulary²⁰. Meanwhile, Coxhead state that academic vocabulary or academic words is a lexical item that occur frequently and uniformly across a wide range of academic material. Townsend define general academic vocabulary as words which are across content areas, have abstract definitions and are challenging to master. Similarly, Hiebert and Lubliner provided the definition: “words whose meanings often change in different content areas”.²¹ While in this research, academic vocabulary means vocabularies that appear across subject matter disciplines but may vary in meaning in consequence the discipline itself.

¹⁹ “What Is a Corpus, What Is Corpus Linguistics?” (n.d.): 3.

²⁰ I. S. P Nation, *Learning Vocabulary in Another Language* (Cambridge; New York: Cambridge University Press, 2001), 303, accessed May 23, 2019, <http://dx.doi.org/10.1017/CBO9781139524759>.

²¹ James F. Baumann and Michael F. Graves, “What Is Academic Vocabulary?,” *Journal of Adolescent & Adult Literacy* 54, no. 1 (2010): 6.

3. Thesis Proposal

Thesis is the product of a scholarly and professional study at the honors or the master's degree level. It is usually a document in a format and style specified by the particular university. Meanwhile, proposal is a written plan for a thesis or for a dissertation developed by a student for consideration and possible approval by a Thesis or Dissertation committee.²²

While in this research, Thesis proposal means one of a subject matter of the seventh semester students in academic year 2018/2019. This subject is the last section of writing skill. The order of subject matter of writing skills in English teacher Education Department of Islamic University of Sunan Ampel Surabaya are Paragraph Writing in, Essay Writing, Argumentative Writing and Research Proposal Writing.

²² James Mauch, *Guide to the Successful Thesis and Dissertation: A Handbook For Students And Faculty, Fifth Edition*, 5th ed. (CRC Press, 2003), 10–11, accessed August 2, 2019, <https://www.taylorfrancis.com/books/9781482293081>.

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter discusses about all of related review of the study. There are four points that discussed in this chapter, those are: corpus analysis, writing, vocabulary, and academic vocabulary. In addition, some previous studies related to academic vocabulary field also will be revealed.

A. Theoretical Foundation

1. Corpus Analysis

a. Definition of Corpus Analysis

Any old collection of text does not make a corpus. It must represent something, and its merits will often be judged on how representative it is. Corpus (plural: corpora) is a collection of texts, written or spoken, which is stored on a computer.²³ Because corpora are stored on a computer, this allows for very large amounts of text to be compiled and analysed using specially designed software.

Linguistic being the scientific study of language and its structure, corpus linguistic or corpus analysis is the study of language based on text corpora also a methodology to obtain and analyse the language data either quantitatively or qualitatively. We can look at a language feature in a corpus in different ways. For example, using corpus of newspapers, we could examine how many times the words *fire* and *blaze* occur. This will give us quantitative results. However, another approach is to look more qualitative at how a word or phrase is used across a corpus. To do this, we need to look beyond the frequency of the word's occurrence. As we will exemplify below, looking at concordance lines can help us do this and to see qualitative patterns of use beyond frequency.

²³ Anne O'Keeffe, Michael McCarthy, and Ronald Carter, "From Corpus to Classroom: Language Use and Language Teaching" (n.d.): 1.

Some of basic techniques or tools that can be used on a corpus analysis described below:

1) Concordancing

Concordancing is a core tool in corpus linguistics and it can find every occurrence of a particular word or phrase using a corpus software. With a computer, we can search millions of words in seconds. The keyword is often referred as the 'node' and concordance lines are usually presented with the node word/phrase in the centre of the line with seven or eight words presented at either side known as Key-Word-In-Context display or KWIC concordance.

2) Word Frequency Counts or Wordlists

Another common corpus technique which software can perform is the extremely rapid calculation of word frequency list for any batch of text. By executing a word frequency list, you can get rank ordering of all the words in it in order of frequency. Frequency is one of the most important characteristics of vocabulary, it affecting most or all aspects of lexical processing and acquisition. Therefore, corpus data is the best source of frequency information and several findings.²⁴ Frequency is the bedrock of corpus linguistics. At its simplest level, frequency refers to the number of times something occurs in a corpus (or text).²⁵

3) Keyword Analysis

This function allows us to identify the key words in one or more texts. Key words are not usually

²⁴ Schmitt, *Researching Vocabulary : A Vocabulary Research Manual*, 13.

²⁵ Paul Baker, "Sociolinguistics and Corpus Linguistics" (n.d.): 19.

the most frequent words in a text or collection of texts, rather they are the more unusual frequent.²⁶

4) Collocation

Collocation is a way of demonstrating exclusive or frequent relationships between words. If two words collocate, then they have a tendency to occur near or next to each other in naturally occurring language use. For example, *drop off*, *swimming pool*, *letter box* and so on. Collocation therefore indicates a relationship, but we may need to carry out concordancing work in order to identify exactly how the relationship is manifested in language.

The use of corpora has covered up in many different aspects, start from crime until about language itself. First is lexicography. Language corpora have many applications beyond language description for its own sake. They are now the standard tool for lexicographers, who use multimillion words corpora to examine word frequency, patterning and semantics in the compilation of dictionaries. The Cambridge International Corpus (CIC) has over one billion words. They are constantly added to and facilitate the monitoring of language trends and usages.

The second is corpora also used in grammar aspects, the COBUILD project, or *Collins Birmingham University International Language Database* was set up at University of Birmingham in 1980 under command of John Sinclair, they had a major influence on grammar and also provide very useful information about the types of lexical and grammatical errors that allows for dictionary writers or other material writers to highlight typical problems.²⁷

²⁶ Anne O'Keeffe, Michael McCarthy, and Ronald Carter, "From Corpus to Classroom: Language Use and Language Teaching" (n.d.): 12.

²⁷ *Ibid.*, 17.

The third aspect is corpus usage another area which is increasingly using language corpora as a tool is forensic linguistics. It concerns with the use of language in law and crime investigation. Corpus or corpora have many applications relative to diversity of the focus of the discipline itself include the analysis of the genuineness of documents from confession to suicide notes, authorship identification in academic settings such as issues of plagiarism, threat letters, forensic phonetics such as speaker identification, police interview and interrogating data.²⁸

For the last aspect, corpora also had an impact in the area of sociolinguistics. Many corpora of spoken language make this area is not surprisingly, in particular can be built around sociolinguistics variables such as age, gender, level of education, socio economic background and so on. Ihalainen has stated in terms of regional variation. He compared the grammatical subject in educated and dialectal English in the London-Lund and the Helsinki Corpus of modern English dialects.²⁹ For other aspects such as age-related research has been done by Haslerud and Stenstrom.³⁰ They focused on context of teenager language called The Bergen Corpus of London Teenager Language (COLT).

²⁸ Lawrence M Solan and Peter M Tiersma, "Author Identification in American Courts," *Applied Linguistics* 25 (n.d.): 448–465.

²⁹ Ossi Ihalainen, Terttu Nevalainen, and Matti Rissanen, "History of Englishes: New Method and Interpretations in Historical Linguistics," in *The Grammatical Subject in Educated and Dialectal English: Comparing the London-Lund Corpus and the Helsinki Corpus of Modern English Dialects*, Topics in English Linguistics 10 (Johansson, Stig; English Computer Corpora: Selected Papers and Research Guide; De Gruyter Mouton: Berlin, Boston, 1992), 799.

³⁰ Vibecke Haslerud and Anna-Brita Stenstrom, "The Bergen Corpus of London Teenager Language (COLT)," *Spoken English on computer: Edited by Geoffrey Leech, Greg Myers, Jenny Thomas* (n.d.).

There are some tools for analysing corpus. First is AntConc,³¹ it is a freeware corpus analysis for concordance and text analysis. Some facilities we can get from this application such as concordance, concordance plot, cluster/N-Grams, collocates, word list, and keyword list.

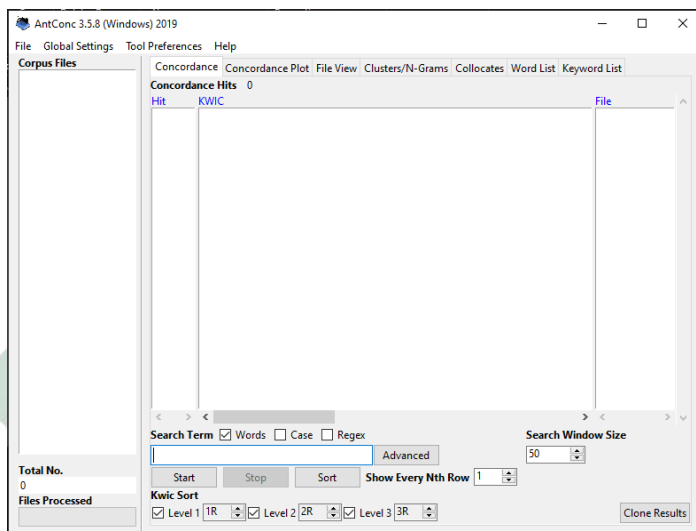


Figure 2.1: AntConc main window

Second is WordSmith³² tools, it is an integrated suite of programs for looking at how words behave in texts. You will be able to use the tools to find out how words are used in your own text. But we have to register first to be able to use WordSmith application entirely, since this is paid application.

³¹ Laurence, "Laurence Anthony's AntConc."

³²"WordSmith Tools," accessed <https://lexically.net/wordsmith/index.html>.

August 20, 2019,

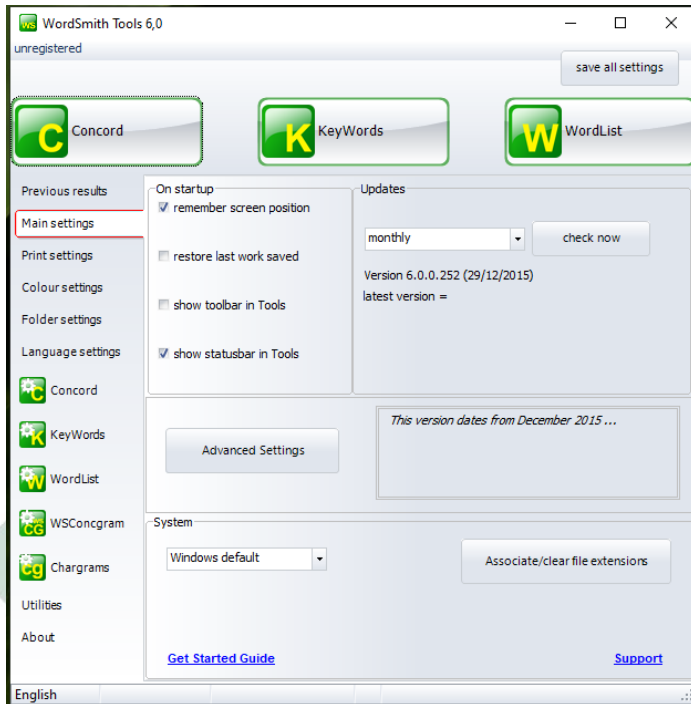


Figure 2.2: WordSmith main window

The last is freeware application called AntWordProfiler,³³ multiplatform tool for carrying out corpus linguistics research on vocabulary profiling. AntWordProfiler contains two tools. The first tool is File Viewer and Editor Tool. This tool allows you to view an individual user file and highlight the different levels of vocabulary in the file using a colour coding. It also shows the overall coverage of different vocabulary levels. The tool allows you to edit the file and immediately see the effect of that editing on the vocabulary level of the words and the text as a whole. If the thesaurus option is activated in the main menu, vocabulary items with thesaurus

³³ Anthony, "Laurence Anthony's AntWordProfiler."

entries will be highlighted and equivalent items at a lower (or higher) vocabulary level can be chosen as replacements.

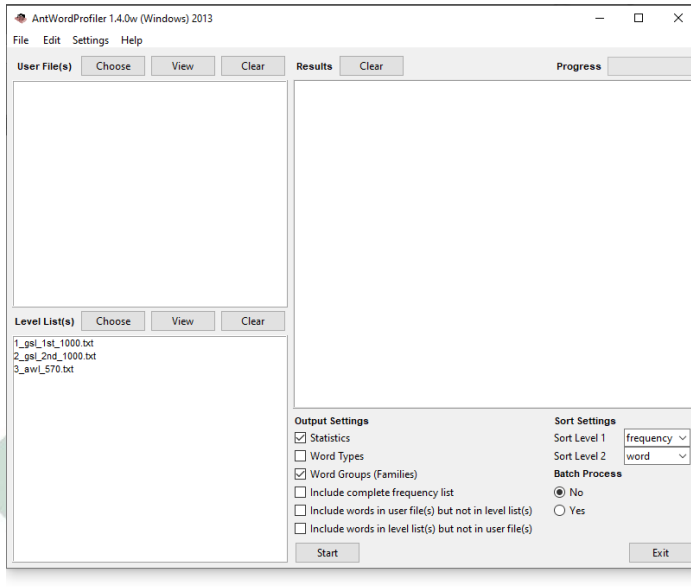


Figure 2.3: AntWordProfiler main window

The second is a general vocabulary profiling tool. It is the main tool and it appears in the main window of the program. This tool allows you to generate statistics and frequency information about corpus of text loaded into the program. It compares the files against a set of vocabulary level lists that can be plain frequency lists of ‘family lists’ based on the research of Paul Nation. We can include any kinds of corpus or word list that can be use as measuring standardization.

Most of corpus analysis software allow users to produce word lists (literally a list containing every word), which give frequencies (both raw and standardised) within a corpus.

Such word lists are normally presented in two ways alphabetically and by frequency.³⁴

b. Types of Corpora

5) General Corpora

The texts that do not belong to a single text type, subject field, or register, may include written or spoken language, or both. They aim to represent language in its broadest sense and to serve as a widely available resource for baseline or comparative studies of general linguistic features. For example, Brown corpus, BNC (British National Corpus).

6) Specialized corpora

a specialized corpus contains texts limited to one or more subject areas, domains, topics etc. such corpus is used to study how the specialized language is used. It aims to be representative of a given type of text and used to investigate a particular type of language. For example: Cambridge and Nottingham Corpus of Discourse in English (CANCODE).

7) Diachronic or Historical Corpora

The texts are usually from different periods of time. The aim is at representing an earlier stage of a language and they help to trace the development of a language overtime. For example: Helsinki Corpus.

8) Learner's Corpora

Aim at representing the language as produced by the learners of a language, and they include spoken or written language samples produced by non-native speakers. The corpus is used to study the mistakes and problems learners have when learning a foreign language. For example: Louvain Corpus of Native

³⁴ Baker, "Sociolinguistics and Corpus Linguistics," 20.

English Essays (LOCNEE), International Corpus of Learner English (ICLE).

9) Parallel Corpora

A parallel corpus consists of two monolingual corpora. One corpus is the translation of the other. Parallel corpus can be used by translator and by learners to find potential equivalent expressions in each language and to investigate differences between languages. For example, a novel and its translator or translation memory of a CAT tool could be used to build a parallel corpus. Both languages need to be aligned, i.e. corresponding segments, usually sentences or paragraphs, need to be matched.

10) Multilingual Corpus

A multilingual corpus is very similar to a parallel corpus. The two terms are often used interchangeably. Multilingual corpus is any systematic collection of empirical language data enabling linguists to carry out analyses of multilingual individuals, multilingual societies or multilingual communication.

11) Comparable Corpus

Two or more corpora in different languages (e.g. English and Spanish) or in different varieties of language (e.g. Indian English and Canadian English). Comparable corpora of different language can be used by translators to identify differences and equivalences in each language. For example: International Corpus of English (ICE).

Furthermore, in this research the researcher uses AWL (Academic Word List) compiled by Averil Coxhead as a Corpus-Analysing tool. Academic Word List has been designed upon an analysis of 3.5 million words corpus of written academic English from 28 sub-disciplines of four main disciplines; Art, Law, Commerce, and Science. The Academic Word List are consists of 570 word families selected in

agreement with three criteria: 1) frequency of occurrence (occurrence at least 100 times in the corpus), 2) range (occur at least 10 times in each of the four disciplines and from 15 or more subject areas), 3) specialized appearance (be outside the first 2000 GSL words).³⁵ The Academic Word List covers up approximately 10% all of the vocabulary covered in a piece of academic writing. This means that, on average, one word in 10 in an academic text book is in the Academic Word List.³⁶

2. Writing

a. Definition of Writing

Writing is an intellectual activity to finding some ideas and how to express and arrange them into statement and paragraph that can understand by the people. In writing, as a writer we are demanded to show the thoughts and how we organize them into a good composition, because writing presents the writer's concept in understanding an issue which is shown to the public.

Writing is a progressive activity which demands intellectual efforts. This means that when we first write something down, we have already been thinking about what we are going to say and how we are going to say it. It involves generating ideas, planning, goal setting, then after we finished writing, we read over what we have written and make changes and corrections. Therefore, writing is never a one-step action to convey the message of writing.

b. Academic Writing

Most academic courses assess students through written assignments. These include coursework, which may take weeks to write, and exam answers, which often have to be written in an hour or less. Writing involves starting, progressing and

³⁵ Jamalzadeh and Chalak, "A Corpus-Based Study of Academic Vocabulary in Physiotherapy Research Articles."

³⁶ Averil Coxhead, "An Introduction to the Academic Word List" (n.d.): 9.

finishing a complicated, challenging combination of tasks. It requires you to active lots of different skills and orientations, sometimes at different stages and phases in the process, sometimes all at the same time. Academic writing is the kind of writing used in high school and college classes. Academic writing is different from creative writing, which is the kind of writing you do when you write letters or e-mails to your friends and family. Creative writing is informal, so you may use slang, abbreviations, and incomplete sentences. However, academic writing is formal, so you should not use slang or contractions. Writers should be clear why they are writing. The most common reasons for writing are; to report on a piece of research the writer has conducted; to answer a question the writer has been given or chosen; to discuss subject of common interest and give the writer's view. Alice Oshima and Ann Hogue³⁷ in their book entitled *Writing Academic English*, they state that there are two types of academic writing. The first is writing a paragraph and the second is writing an essay.

1) Paragraph writing

A paragraph is a group of related sentences that discuss one (and usually only one) main idea. A paragraph can be as short as one sentence or as long as ten sentences. The number of sentences is unimportant; however, the paragraph should be long enough to develop the main idea clearly.³⁸ There are three part of paragraph, all paragraphs have a topic sentence and supporting sentence, and some paragraph also have a concluding sentence.

- a) Topic sentence states the main idea of the paragraph and the most general statement in the paragraph because it gives only the main

³⁷ "Writing Academic English, 4th Edition.Pdf," n.d.

³⁸ *Ibid.*, 2.

idea, it does not give any specific details. It briefly indicates what the paragraph is going to discuss. Besides that, topic sentence not only mentions the topic, but also limits the topic to one specific area that can be discussed completely in a single paragraph.

- b) Supporting sentences are to develop the topic sentence. They explain or prove the topic sentence by giving more information and example about it.
- c) Concluding sentence is signals the end of the paragraph and leaves the reader with important points to remember.

2) Essay Writing

An essay is a piece of writing several paragraphs long. However, because the topic of an essay is too complex to discuss in one paragraph, we need to divide it into several paragraphs, one for each major point. Writing an essay is no more difficult than writing a paragraph except that an essay is longer. The principles of organization are the same for both, so if we can write a good paragraph, we can write a good essay.³⁹ There are three part of an essay, they are introduction, body and a conclusion.

³⁹ Ibid., 56.

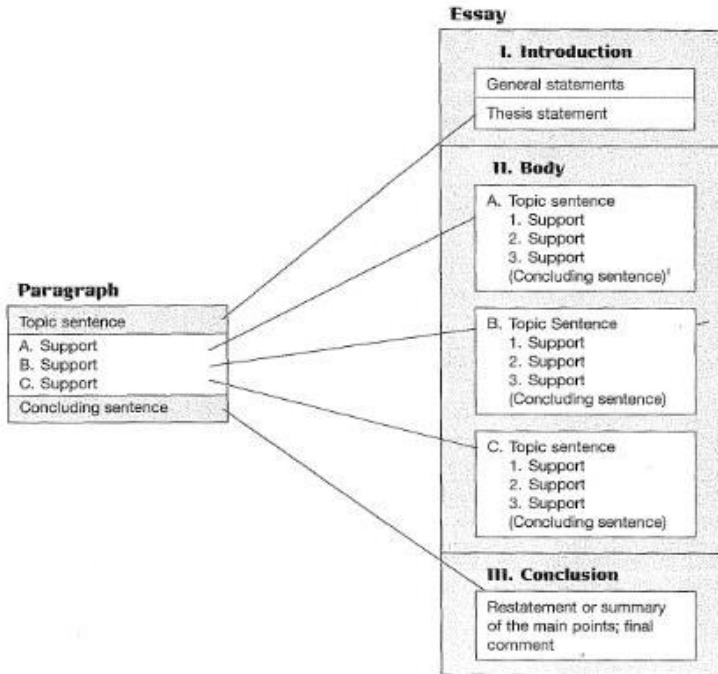


Figure 2.4 Part of Essay

- 1) Introduction consists of two parts: a few general statements to attract the reader's attention and a thesis statement to mention the main idea of the essay. It names the specific topic and gives the reader a general main idea of the contents of the essay.
- 2) The body consists of one or more paragraphs. Each paragraph develops a subdivision of the topic, so the number of paragraphs in the body will be vary with the number of subdivisions or subtopics.
- 3) The conclusion, like the concluding sentence in a paragraph, is a summary or review of the main point discussed in the body.

There are more types of academic writing based on their features and purposes. They are descriptive, analytical, persuasive and critical.

1) Descriptive

Descriptive is the simplest type of academic writing. Its purpose is to provide information or facts. The example would be a summary of an article or a report of the result of an experiment.

The assignment with purely descriptive instruction includes identify, report, record, summarise and define.

2) Analytical

For university-level text it is rare to use purely descriptive, most academic writing is also analytical. Analytical writing includes descriptive writing, but we have to re-organise the information and the fact we described into categories, parts, types, groups or relationships.

The kinds of instructions for analytical assignment includes analyse, relate, contrast, compare, examine.

There are some ways to make our writing more analytical. Firstly, we have to spend plenty of time for planning, brainstorm the facts and ideas, and try different ways of grouping them, according to patterns, similarities, parts, differences. We also can use flow charts, colour coding, or tables. Secondly, build each section and paragraph around one of analytical categories. For the last, we have to make the structure of our paper clear to our reader by using sentences and clear introduction.

3) Persuasive

Persuasive type of writing has all the features of analytical writing includes information and re-

organise the information. But in persuasive we add our point of view. Most essays are persuasive, and there is a persuasive element in the discussion and conclusion of research article.

Point of view in academic writing can consist of an argument, a recommendation, interpretation of findings or evaluation of the work of others. In persuasive, each statement you make needs to be supported by some evidence. The kinds of persuasive assignment include evaluate, argue, discuss.

There are some ways to help reach our point of view on the facts or ideas. Firstly, we have to read some other researchers' point of view on the topic. Secondly, look for patterns in the data or references, where is the strongest evidence. For the last, we have to discuss the facts and ideas with someone else to make sure the point of view is subtle and appropriate enough.

To develop and present our argument, there are some ways to help. Firstly, we have to list the different reasons of our point of view. Secondly, try to think about different types and source of evidence which can support our point of view. The third is our text develops a coherent argument to support our point of view. Beside that we have some convincing and relevant evidence for every statement we made.

4) Critical

Critical has all feature of persuasive writing with added feature of at least one other point of view. While persuasive writing requires our own point of view, critical writing require at least two points of view including ours.

Example of critical writing assignment include a critique of a journal article, or a literature review that identifies the strengths and weakness of existing research.

Critical writing requires strong writing skills. We need to understand the whole topic and issues. We need to develop an essay structure and paragraph structure that allows us to analyse different interpretations and develop our own argument supported by evidence.⁴⁰

3. Vocabulary

a. Definition of Vocabulary

Vocabulary is the series of words that make up a language. It includes where the word comes from, how they change, how they relate to each other and how to use them to view the world. Vocabulary is an important part of English language teaching because without adequate vocabulary teacher cannot deliver or express their own idea about the material and students cannot understand well about what teacher explain to them. To make our students satisfied with our explanation, we as a good teacher should keep increasing our capability in acquiring and using proper vocabulary when delivering the material. After they find themselves having more understanding about what they read and what they hear, it will be easier for them to express themselves because they have a better command of their language. therefore, vocabulary can be an indication of how well a learner can comprehend the language.

The most ambitious goal is to know all of vocabulary or word of the language. This is very ambitious because native speakers itself doesn't know all the vocabulary of the language because there are numerous specialist vocabularies, such as the vocabulary of nuclear physics or computational linguistics, which known by small groups of people who specialize in those areas.

⁴⁰ "Types of Academic Writing," *The University of Sydney*, accessed February 12, 2019, /students/writing/types-of-academic-writing.html.

b. Types of Vocabulary

Differentiating the types of vocabulary can help teachers to decide how much effort to put into teaching particular words, according to Nation, vocabulary divided into three types, they are:

1) High-Frequency Word

A high frequency word is a small group of words which are very important because these words cover a very large proportion of the running words in spoken and written texts. The most commonly used list of high-frequency words is the General Service List of English Words (GSL) by Michael West which includes around 2.000 words families. About 165 words families in this list function such as a, some, two, to and because. The rest are content words; nouns, verbs, adjective, and adverbs, for example: government, introduce and popular. High-frequency cover almost 80% of the running words in the academic text and it will give a lexical coverage of around 85% in written text and 90% in unscripted spoken discourse.⁴¹

2) Low-Frequency Word

In the text, they are marked italics. They are the biggest groups but only cover about 5% of the running words in an academic text. They are including:

- a) Proper noun. It refers to a word which is used for a person, place or organization. A proper noun always begins with a capital letter no matter where it occurs in a sentence, for example: London, John, Australia.
- b) Words of moderate frequency that did not manage to get into high frequency lest.
- c) Some low-frequency words are simply low-frequency words which is the words are

⁴¹ I. S. P. Nation, *Learning Vocabulary in Another Language* (Cambridge: Cambridge University Press, 2000), 21.

almost rarely used by the language user, for example: gibbous, plummet, bifurcate.

3) Specialized Vocabulary

Specialized vocabulary is the words used for talking about the particular subject because specialized vocabularies are made by systematically restricting the range of topics or language uses investigated.⁴² That is the possible reason to have special vocabularies for any occasion for example speaking, reading academic text, for reading newspaper, for reading children's stories or may be for letter writing. These words are a way of extending the high-frequency word for special purposes. There are two kinds of specialized vocabulary, they are; Academic words and Technical words.⁴³

4. Vocabulary Usage

Cultural variation in the function and uses of language has important consequences for speakers of variants, especially with respect to educational performance. There are three consequences can be explained: social, cognitive and acquisition school skills.

The social consequences of a variant way of using language can affect teacher-pupil as well as peer relationships. The consequences of a teacher's attitude towards a given dialect are thoughtful. For example, it can be affecting his or her judgment about how a smart child is likely to be, how he or she will be grouped and how his or her contribution in class will be treated. This, in turn can affect the child's attitude about himself as a school learner, his willingness to participate, his expectations about result of his participation.

⁴² Nation, *Learning Vocabulary in Another Language*, 24.

⁴³ I. S. P. Nation, *Learning Vocabulary in another Language...28 - 29*

Also, an issue in this present work is whether different patterns of language socialization in the home, in this case, vocabulary use. It has obvious cognitive consequences. Vocabulary differences absolutely reflect differences about one's ideas in public communication. These differences lead to different opportunities to talk about a given meaning or aspect of meaning, as a result of a different speech communities have different access to its members' and others' ideas. For example, the process of modification in the case of adjectives or adverbs or the process of subordination in the case of conjunctions could easily be affected by differentially elaborated vocabularies.

The possible consequences of variants from the school register for the acquisition of school skills may be illustrated for reading and the ability to deal with a kind of meta-behavioral information. In reading, semantic mismatch between reader's word meaning and author's meaning may affect children's expectations about the idea of the language that they are reading. Variation in language socialization may also differentially facilitate or support the child's growing ability to analyse and make analytical statements about kinds of behaviour which are not happen in everyday life.⁴⁴

5. Academic Vocabulary

a. Definition of Academic Vocabulary

According to Nation, Academic vocabulary includes into specialized vocabulary called Academic Word List. The researcher uses Academic Word List by Averil Coxhead as a standard measurement in this study, because this academic word list was selected by examining a large corpus (or collection) of written academic texts and selecting the words that occurred: 1) in text from all four academic faculty section: Arts, Commerce, Law, and

⁴⁴ Christine Coombe, "Assessing Vocabulary in the Language Classroom" (n.d.): 11–12.

Science. 2) over 100 times in the corpus overall. 3) at least 10 times in each academic faculty section. 4) outside the 2000 most frequent words on Michael West's *General Service List* (GSL) the GSL includes everyday words such as *I, House* and *Do*.⁴⁵

The Academic Word List is intended as a reference for students who are studying or preparing to study at a tertiary level in English. The Academic Word List does not include 'content' vocabulary for particular subjects which students obviously need to learn as well. The Academic Word List focuses instead on the non-subject-specific vocabulary that students of any discipline will need to master in order to produce coherently-structured written assignments.

The academic vocabulary knowledge is crucial for students studying at an English-speaking university. Academic writing tends to be quite formal in style the vocabulary items that we see in our writing can affect the level of formality. Academic vocabulary helps students understand oral direction and classroom instruction as well as comprehend text across different subject areas.⁴⁶

Following the vocabulary development work of Beck, McKeown and Kucan, the CCSS (Common Core State Standard) references that are vital to academic achievement fall into three tiers:

1. Tier one words include basic or high-frequency vocabulary words and usually don't include multiple meaning words. Words are the words of everyday speech usually learned in the early grade.
2. Tier two words are less familiar to students but help in comprehending written text and conversation between the teacher and student. They appear in all sorts of texts: informational texts (words such as *relative, vary,*

⁴⁵ Coxhead, "An Introduction to the Academic Word List."

⁴⁶ Ibid.

formulate, specificity, and accumulate), technical texts (*calibrate, itemize, periphery*), and literary texts (*dignified, faltered*). The words are precise but more subtle forms of familiar words and include multiple meaning words.

3. Tier three words are “domain specific” and are critical understanding subject content. Generally, they have low frequency use and are limited to specific subject. For example, in the geographical terms’ *peninsula*, and *cape*.⁴⁷

Based on the explanation above, academic vocabulary is categorized in tier two words because academic vocabulary can be defined as that are not tied to any particular course and its meanings often change in different content areas and include multiple meaning words. Academic vocabulary words often show up as the skill words in learning objective but can occur at any point during a lesson as they arise within the context of a lesson. For example: *distribute, establish, occur, and recognize*.

It has contrast differentiation between academic vocabulary and content vocabulary. For contrast vocabulary, include words that are specific to a given domain or subject area. Often, these words can be found directly in the content standards and are presents as the concept words (typically noun) in learning objective. Based on the explanation above, content vocabulary is categorized in tier three word. For example: *affixes, exponents, and photosynthesis*.⁴⁸

⁴⁷ “BUSD Grade Level Academic Vocabulary,” n.d., 2.

⁴⁸ DataWORKS Curriculum, “Vocabulary Development Part 1: Types of Vocabulary,” *Dataworks Educational Research*, July 29, 2014, accessed August 10, 2019, <https://dataworks-ed.com/blog/2014/07/vocabulary-development-part-1-types-of-vocabulary/>.

Types of Vocabulary (Across Grades)

DataWORKS		Academic Vocabulary • used across all disciplines <i>(Often not taught in Textbooks)</i> <i>Examples:</i> <i>distinguish, corresponds, combine, separate, analysis, symbolic</i>	Content Vocabulary • content specific <i>(Taught during Concept Development in EDI Lessons)</i> <i>Examples:</i> <i>main idea, thesis statement, figurative language, denominator, linear equation, addition, ratios, perimeter, Civil War, separation of powers, legislative branch, mitosis, cell wall, photosynthesis, Solar System</i>	Support Vocabulary • in specific textbooks and worksheets; may be challenging for EL students <i>(Often over-emphasized in Textbooks)</i> <i>Examples:</i> <i>hailout, hammock, port, starboard</i>
	Common Core	Tier One words <i>(everyday speech)</i> <i>Beginning ELD</i>	Tier Two words <i>(general academic words)</i> <i>Examples in Informational text:</i> <i>relative, vary, formulate, specificity, accumulate</i> <i>Examples in Technical text:</i> <i>calibrate, itemize, periphery</i> <i>Examples in Literary text:</i> <i>misfortune, dignified, faltered, unabashedly</i>	Tier Three words <i>(domain-specific words)</i> <i>Examples:</i> <i>lava, legislature, circumference, aorta</i>

Figure 2.5: Types of Vocabulary

Academic vocabulary accounts for a substantial number of words in academic text. There are two ways of measuring academic vocabulary. First is by looking at the number of tokens (coverage) academic vocabulary accounts for, secondly by looking at the number of types, lemmas or word families. Therefore, it is important to understand the ways to count the words. According to Nation there are several ways of counting words, that is deciding what will be counted.

1. Tokens

Tokens are total number of words. Every word form in sentence is counted although the same word form that occurs more than once. For example, the sentence “it will not easy to pronounce it correctly” it contains eight tokens.

2. Types

In this way of counting words, if there is a word which occurs more than once then it will not be counted again. So, the sentence of eight tokens “it

will not easy to pronounce it correctly” consists of seven types.

3. Lemmas

Lemmas are root forms in each word class. It consists of a headword and some of its inflected forms. The English inflections consist of plural, presents tense, past tense, past participle, present participle, comparative, superlative, and possessive. For example, [verb: jump, jumps, jumped, jumping], [noun: jump, jumps], [noun: jumper, jumpers] consist of 3 lemmas, 6 types, 8 tokens.

Using the lemma as the unit of counting greatly reduces the number of units in a corpus. Bauer and Nation calculate that the 61.805 tagged types (or 45.957 untagged types) in the Brown corpus become 37.617 lemmas which is a reduction of almost 40% (or 18% for untagged types).

4. Word family

Word families are root forms. All regularly inflected and derived words sharing a root form. The meaning of the derived word must be closely related to the meaning of the base word. For example, [kick, kicks, kicked, kicking, kicker, kickers], [do, doing, undo, redo, doable] consist of 2 word families, 7 lemmas, 11 types, 11 tokens.⁴⁹ In order to measure vocabulary profile by AntWordProfiler, token percentage are used as the method of counting words.

A research has been conducted by Sutarsyah, Nation and Kennedy,⁵⁰ they found that academic vocabulary (the University Word List) accounted for 8.4% of the tokens in the learner and scientific sections (Section J) of the *LOB* and

⁴⁹ Nation, *Learning Vocabulary in Another Language*, 10–11.

⁵⁰ “Sutarsyah, Cucu & Nation, Paul & Kennedy, Graeme. (1994). How Useful Is Eap Vocabulary for Esp? A Corpus Based Case Study. *Relc Journal*. 25. 34-50. 10.1177/003368829402500203.,” n.d.

Wellington Corpora, and 8.7% of the tokens in an Economic text.

In this study, the researcher did the corpus analysis of thesis proposal to examine how many times the academic vocabulary occurs in each thesis proposal using Academic Vocabulary List (AWL). With this vocabulary list we can examine how many times the word such as *research* and *definition* occur and it will give us a quantitative data.

B. Previous Study

The researcher found some similar topics of the previous study. The study also discusses about proposal writing. The first previous study is the research entitled “Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups” By Habibullah Pathan, Rafique A. Memon, Shumaila Memon, Syed Waqar Ali Shah, and Aziz Magsi. This research analyses the frequency of the AWL Word Families in the Pakistani doctoral theses text. The source of the data comprises of 200 doctoral theses from two major scientific disciplinary groups. They use concordance software AntConc by Anthony Laurence. The result of the study revealed that in 8.76% (496839 words) of the Pakistani doctoral theses is covered by the AWL words. they make attempt further analyse the AWL text coverage by considering the frequency of occurrences in terms of word families. The finding showed that among 570 word families of AWL, 550 word families with the sum of 96.49% are found to occur more than 10 times in Pakistani doctoral theses.

The second previous study is the research entitled “Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners’ English”⁵¹ by Batia Laufer and Tina Waldman. This research is to investigate the use of English verb-noun collocations in the writing of native speakers of Hebrew at three proficiency levels. They compiled a learner corpus that consists of about 300,000

⁵¹ Batia Laufer and Tina Waldman, “Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners’ English: Verb-Noun Collocations in L2 Writing,” *Language Learning* 61, no. 2 (June 2011): 647–672.

words of argumentative and descriptive essays. For comparison purposes, they select LOCNEE (The Louvain Corpus of Native English Essays). They retrieved the 220 most frequently occurring nouns in the LOCNEE corpus and in the learner corpus, created concordances for them, and extracted verb-noun collocations. Subsequently, they performed two types of comparisons: learners were compared with native speakers on the frequency of collocation use and learners were compared with other learners of different second-language proficiencies on the frequency and correctness of collocations. The data show that learners at all three proficiency levels produced far fewer collocations than native speakers, that the number of collocations increased only at the advanced level.

The third previous study is the research entitled “Academic Words in Education Research Articles: A corpus study” by Ali Mozaffari and Raouf Moini from English Department of University of Kashan, Iran.⁵² This research conducted to investigate the importance and relevance of general academic vocabulary list in a range of different disciplines. 1.7 million words corpus was compiled consisting of 239 research articles published by *Educational Research Review*, *International Journal of Educational Development*, *International Journal of Educational Research*, *Linguistic and Education*, *Teaching and Teacher Education* during period of 2002 to 2010. They analyse the data using computer software *WordSmith*. Computer analysis showed that the AWL words accounted for 4.94% (84115 of the running words). An attempt was also made to identify the most frequent word families of AWL in the Education Research Articles Corpus as well as to obtain a list of non-AWL words (not include in Coxhead’s list nor West’s General Service List).

The fourth previous study was conducted by Min Zhang from Beijing Information Science and Technology University where her interest includes corpus linguistics, computer-assisted language learning and academic reading and writing. The quantitative analysis research entitled “A Corpus-based Comparative Study of Semi-technical and Technical Vocabulary” aimed at comparing semi-

⁵² Ali Mozaffari and Raouf Moini, “Academic Words in Education Research Articles: A Corpus Study,” *Procedia - Social and Behavioral Sciences* 98 (May 2014): 1290–1296.

technical vocabulary and technical vocabulary to address the specific needs of undergraduates majoring in information engineering in mainland China. A 1,024,882 words corpus of Information Engineering English Corpus (IEEC) was built using texts from ten specialized courses. Semi-technical and technical vocabulary items were profiled using West's General Service List and Coxhead's Academic Word List. As a result, 248 semi-technical and 166 technical word families were extracted covering 9.16% and 4.95% of the total tokens of the corpus.

The last previous study is the research entitled "Comparison Analysis on AWL of the Articles in Indonesian Journal of Applied Linguistic and Oxford University Press of Applied Linguistic" by Mahsyurotun Nikmah. This study focuses on comparing the utilization of Academic Word List (AWL) from the local and international journal articles. The articles that in this study must meet two criteria. First, the articles should have open access articles. Eventually it is found the journal articles from Indonesian Journal of Applied Linguistic (IJAL) and Oxford University Press of Applied Linguistic (OUP) that provide the open access articles. Second, the articles should have newest issues that have been published in the last two years.

The result of the study is 507 academic word families of Academic Word List from IJAL and 535 academic word list of OUPs were found. The researcher state that there are three steps to analyse the Academic Word List of the journal articles, comprising the percentage, the differences and similarities and also the distribution of sub list Academic Word List. The articles of IJAL have percentage of Academic Word List about 10.22% but the OUP article shows 11.41%. Also, there are 68 different academic words from the articles of IJAL and OUP. The 20 words belong to IJAL and the other 48 words belong to OUP. It can be concluded that both of the journal articles of IJAL and OUP are categorized as academic journal.

CHAPTER III

RESEARCH METHODOLOGY

This chapter discusses about: (a) approach and research design, (b) subject and setting, (c) data and source of data, (d) data collection technique, (e) research instrument, (f) data analysis technique, (g) research stages.

A. Approach and Research Design

The research design of this study is quantitative descriptive method to find out the optimal frequency threshold for academic vocabularies. Also, the researcher needs the numeric and descriptive data. Since Creswell state about some characteristic of quantitative research that is collecting numeric data from a large number of people using instruments with pre-set questions and responses.⁵³ This study conducted to analyse the use of academic vocabulary in composing thesis proposal writing by seventh semester students in academic year 2018/2019.

B. Subject of the Study and Setting

1. Population

A population is a group of individuals who have the same characteristic.⁵⁴ In this study, the population is 113 seventh semester students in academic year 2018/2019 of English Teacher Education Department in UIN Sunan Ampel Surabaya who enrol in Thesis Proposal Seminar class. It is divided into 5 classes they are A class (23), B class (22), C class (23), E class (22), and F class (23).

2. Sample

In quantitative sampling strategies, there are two types of strategies those are Probability sampling which the researcher selects individuals from the population who are representative

⁵³ John W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 4th ed. (Boston: Pearson, 2012), 13.

⁵⁴ *Ibid.*, 142.

of that population and Nonprobability sampling where the researcher selects individuals because they are available.⁵⁵ In this study, the researcher used probability sampling and specifically at simple random sampling to select the sample. In random sampling the researcher selects participants (or units, such as schools) for the sample so that any individual has an equal probability of being selected from population.⁵⁶ The intent of simple random sampling is to choose individuals to be sampled who will be representative of the population. Based on lecturer's advice the researcher took 3 proposal documents randomly as a representative from each class. In total the participants are 15 students.

C. Setting of the Study

The location of the study is in thesis proposal seminar class in English Teacher Education Department of state Islamic university of Sunan Ampel Surabaya. This class is considerate become a setting of this study because in this course students practice about how to write a thesis proposal appropriately and they present it in turns.

D. Data and Source of Data

This study is conducted to analyse the use of academic vocabulary in thesis proposal writing by seventh semester students of English Teacher Education Department of UIN of Sunan Ampel Surabaya. For the purpose of the study, the data that need to be collected is documentation. To answer the question of this study, students' thesis proposal documents need to be analysed because the researcher needs to know about academic vocabulary use in their proposal writing. During seventh semester, students must write a thesis proposal as an assignment and to be presented.

E. Data Collection Technique

For the students' proposal writing, the lecturer will give the final assignment for the students to write a thesis proposal

⁵⁵ Ibid.

⁵⁶ John W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 4th ed. (Boston: Pearson, 2012), 143.

and present it in the classroom as the fulfilment of the class. After the thesis proposal has been presented, the researcher asked some of the student's thesis proposal to the lecturer.

F. Research Instrument

For supporting data collection technique, the researcher uses an instrument for it. The instrument is Academic Word List (AWL) compiled by Averil Coxhead as a standardization in this study.

G. Data Analysis Technique

In this study, the researcher uses AntWordProfiler application by Anthony Laurence⁵⁷ to measure the documents which included Academic Word List⁵⁸ as the main word list to analyse the use of academic vocabularies in students' thesis proposal writing. In this case, the researcher input all of data into AntWordProfiler application and measure it one by one exhaustively. Data analysis technique explain as follow:

1. The researcher converts the document into plain text form.
2. Researcher input the plain text document into User File box of the application.
3. The researcher input the academic word list into Level List box. This Level List box used to place a standardization files to measure the document, for example General Service List (GSL) compiled by Michael West which contains 2000 word families, and then Academic Word List by Averil Coxhead which contains 570 word families based on a 3.500.000 token corpus of academic English which is divided into four groupings of Arts, Science, Law and Commerce.
4. Press start then wait for application to calculate the data.
5. The result of the application consists of token, percentage of token, type, percentage of type,

⁵⁷ Anthony, "Laurence Anthony's AntWordProfiler."

⁵⁸ Coxhead, "An Introduction to the Academic Word List."

group (families), percentage of group (families) and list of analysed vocabularies from the document.

NAME	FILE	TOKEN	TOKEN%	CUMTOKEN%	TYPE	TYPE%	CUMTYPE%	GROUP	GROUP%	CUMGROUP%
PARTICIPANT A	Gsl 1 st 1000.txt	3147	63.85	63.85	408	46.42	46.42	294	42.30	42.3
	Gsl 2 nd 1000.txt	295	5.96	69.83	78	8.87	55.29	56	8.06	50.36
	Awl 570.txt	620	12.58	82.41	174	19.80	75.09	126	18.13	68.49

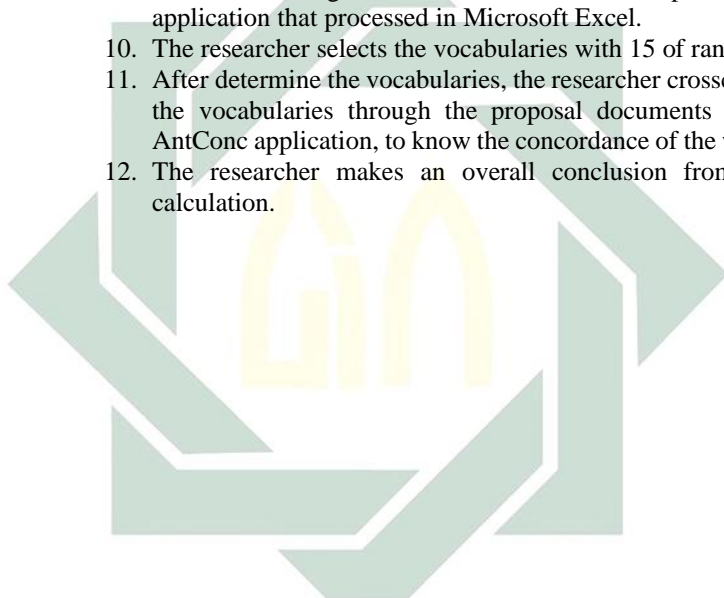
Table 3.1: Example of Result of AntWordProfiler

From this result the researcher obtains a raw data and then it will be interpreted as clear as possible. The aspect that researcher needs is percentage of academic word list (AWL_570) token to know the amount of academic vocabulary used by the students.

H. Research Stages

1. Contact the lecturers of thesis proposal seminar class to ask the thesis proposal that have been presented.
2. The researcher also asks the lecturers about recommendation of the thesis proposal that can be as a sample for this research.
3. The researcher collects the student's document whether a recommendation from the lecturers and for the rest the researcher asks randomly.
4. The researcher converts the document from Microsoft Word extension (.docx) into plain text (.txt).

5. The researcher analyses the documents using AntWordProfiler.
6. The researcher put the thesis proposal documents into User File in main window of the application.
7. The researcher give checklist to some categorizes that need to be analysed
8. The researcher includes the Academic Word List as a standardized measurement in to Level List.
9. The researcher gains a raw data form the output of the application that processed in Microsoft Excel.
10. The researcher selects the vocabularies with 15 of range
11. After determine the vocabularies, the researcher crosscheck the vocabularies through the proposal documents using AntConc application, to know the concordance of the word.
12. The researcher makes an overall conclusion from the calculation.



CHAPTER IV

RESEARCH FINDINGS AND DISCUSSION

This chapter presents the results of the study that were conducted based on the method described in chapter 3. The first section presents the findings of the study about students' academic vocabulary used in their proposal writing. Meanwhile, the second section presents the discussion and implications of the findings on teaching and learning process.

A. Research Findings

This section presents the data findings related to the research question; How is the academic vocabulary used of seventh semester students in their proposal? The findings include: the students' academic vocabulary used in students' proposal writing, and second, data analysis. The total amount of seventh semester students' in academic year 2018/2019 of English language education department of UIN Sunan Ampel Surabaya who enrol in thesis proposal seminar class is 113 divided into 5 classes A, B, C, E, and F. Yet, only 15 documents were taken as the sample randomly and analysed using an application AntWordProfiler.

1. Academic Vocabulary Used in Proposal Writing

The table below describes the top 5 range data of students' academic vocabulary. For the details of the students result, see the table in *appendix 1*.

The Result of Calculation

Vocab	Range	Freq	Par 1	Par 2	Par 3	Par 4	Par 5	Par 6	Par 7	Par 8	Par 9	Par 10	Par 11	Par 12	Par 13	Par 14	Par 15
Analysis	15	118	3	6	26	8	3	16	9	4	12	1	2	8	3	5	12
Data	15	460	41	42	26	39	17	23	35	29	69	10	22	42	43	11	11
Definition	15	31	1	1	2	5	1	2	2	1	5	1	4	3	1	1	1
Method	15	85	9	9	6	7	2	2	14	8	9	6	3	3	1	3	3
Process	15	147	8	16	3	6	11	9	10	14	26	4	28	5	1	3	3
Research	15	694	69	80	43	34	26	39	54	49	64	18	17	85	58	31	27
Researcher	15	414	41	71	26	16	22	28	19	39	77	5	4	22	23	6	15
Scope	15	27	1	3	2	2	2	1	2	1	2	2	1	4	2	1	1
Source	15	50	9	3	3	5	2	4	2	2	5	1	5	3	4	1	1
Technique	15	62	5	6	3	5	2	4	4	3	4	5	2	8	5	4	2
Academic	14	80	2	3	1	1	2	2	5	3	-	2	1	12	16	29	1
Approach	14	44	1	2	-	1	2	2	9	2	3	1	1	5	4	1	10
Conducted	14	52	2	2	-	5	1	5	10	5	10	1	2	4	1	3	1
Focus	14	54	5	11	1	5	2	3	3	4	8	2	3	3	1	0	3
Framework	14	25	2	5	1	1	-	1	1	2	1	1	1	5	1	1	2
Previous	14	114	10	11	8	5	3	9	12	12	14	-	4	2	6	7	11
Significance	14	27	4	1	1	2	3	2	1	2	3	1	1	2	2	2	-
Design	13	84	7	3	7	-	3	3	3	5	2	-	11	22	8	6	4
Journal	13	70	10	4	2	4	6	-	-	5	9	2	5	5	4	13	1
Qualitative	13	125	7	19	16	11	3	1	5	6	12	-	4	17	15	9	-
Analyze	12	51	11	1	12	8	1	2	2	-	3	3	-	3	2	3	-
Analyzed	12	20	-	1	-	1	1	5	-	1	1	1	1	1	3	3	1
Analyzing	12	34	5	-	5	2	2	2	3	1	2	-	2	5	2	3	-
Conduct	12	22	1	2	2	1	2	-	1	-	2	-	1	2	5	2	1
Role	12	33	1	7	-	1	-	-	1	3	2	1	2	6	5	2	2
Theoretical	12	21	3	5	1	-	-	1	1	1	-	1	1	2	1	1	3

Theory	12	50	2	2	1	9	1	-	4	1	-	-	2	17	1	4	6
Communication	11	55	-	4	1	2	-	4	-	27	3	5	3	2	-	3	1
Defined	11	30	5	-	-	5	1	1	2	2	7	3	-	2	1	1	-
Focuses	11	32	11	1	-	-	1	2	-	7	1	2	1	3	2	-	1
Furthermore	11	33	3	11	2	3	1	-	-	3	-	-	1	1	5	2	1
Grade	11	62	13	5	25	-	-	3	2	1	5	1	4	-	-	2	1
Identify	11	15	-	1	1	3	1	3	-	1	1	1	-	1	1	1	-
Participants	11	34	-	10	2	1	3	1	2	-	6	-	2	1	2	4	-
Specific	11	37	4	7	-	1	-	2	4	2	3	1	-	1	2	-	10
Thesis	11	67	5	7	2	-	-	3	2	11	12	3	-	6	6	10	-
Topic	11	28	1	-	-	1	1	1	4	-	3	1	3	5	7	-	1
Appropriate	10	34	1	3	3	-	-	-	6	6	1	2	7	-	1	-	4
consist	10	17	-	-	2	-	1	1	-	1	1	2	1	3	1	4	-
Focused	10	19	-	5	-	1	-	4	1	1	2	1	1	2	1	-	-
Individual	10	34	8	8	-	-	1	1	2	3	5	-	2	2	-	2	-
Methods	10	33	-	3	4	-	3	-	4	-	5	3	1	5	4	1	-
Objective	10	19	2	-	-	2	1	2	-	2	2	-	1	2	1	-	4
Obtain	10	15	1	2	2	-	0	1	1	-	3	2	1	-	1	1	-
Researchers	10	27	2	1	6	4	0	2	1	2	-	-	-	4	2	3	-
Similar	10	13	-	-	-	-	1	1	1	2	1	-	1	2	1	1	2
Vol	10	39	2	-	2	8	6	1	-	-	4	2	-	1	8	5	-

Table 4.1: The Result of Calculation

From table above, the result of AntWordProfiler with explanation as follow: range is the amount of file or files where the current vocabulary exists. Then, frequency is how many times vocabulary comes up. The researcher just sorts the vocabulary according to the top 5 range and present some vocabularies with their spreading all over the participants. For example, word *Analysis* is having 15 on range, means that word *Analysis* are appear on all of the documents, word *Analysis* also have 118 in frequency, means that word *Analysis* appear 118 times through all of the documents and spread into 15 participants. In addition, word *Appropriate* is having 10 on range, means that word *Appropriate* are appear only on 10 documents, word *Appropriate* also have 34 in frequency, means that word *Appropriate* appear 34 times and spread specifically on 10 documents.

To generalize the data, the researcher examines the academic vocabulary that used by all of the participants as shown on the table below:

Number	Vocabularies	Range	Frequency
1	Research	15	694
2	Data	15	460
3	Researcher	15	414
4	Process	15	147
5	Analysis	15	118
6	Method	15	85
7	Technique	15	62
8	Source	15	50
9	Definition	15	31
10	Scope	15	27

Table 4.2: Vocabulary Used by All Participant

In this research, the researcher crosscheck how is the vocabulary used by seventh semester with their proposal document. The researcher only analyzes the context of the vocabularies which have 15 range based on their word class. The researcher only mentions several word classes, because most have the same word class. Here are the details as following:

a. Analysis

Word *Analysis* have 118 frequency, most of them in form of Noun Phrase. For example, sentence “Data analysis is conducted...” by Participant 9 is Noun Phrase. Word *Analysis* in that sentence is as a modifier of *Data*. Then, sentence “The analysis more focused on teacher’s challenge...” by Participant 2 is Noun Phrase with. The last is sentence “Analysis errors means that the researcher will analysis the mistake...” by Participant 6 is Adjective Clause, but the word *Analysis* after *will* in that sentence is wrong, it should be replaced with “*Analyse*” because after auxiliary verb “will” should be bare infinitive.

b. Data

Word *Data* have 460 frequency. First sentence is “There are primary data and secondary data...” by Participant 8 is Noun Phrase with *Data* as the head and *Primary* as the modifier. The last sentence is “The data collecting method used in this research...” by Participant 10 is Noun Phrase with *method* as a modifier, but the word *Collecting* in that sentence is wrong, it should be replaced with “*Collection*” because Noun Phrase did not need Adjective.

c. Method

Word *Method* have 85 frequency. First sentence is “It is a method which people can use...” by Participant 10 is Noun Clause. The last sentence is “The basic step of qualitative method is organizing, focusing...” by Participant 12 is Noun Phrase with *method* as a head and *qualitative* as modifier.

d. Definition

Word definition have 31 frequency. First sentence by Participant 3 is “Definition of Key Term”, it is Noun Phrase with *Definition* as a head. Next is by Participant 11’s sentence “from the definition above, it can be concluded...” it is adverbial phrase with *definition* as head.

e. Process

Word *Process* have 147 frequency, most of them are noun. First sentence by Participant 4 is “to the students in the teaching and learning process...” it is Noun Phrase with process as a head. Next sentence by Participant 6 is “Harmer supposes that grammar is a process for making a writer’s meaning clearer...”, it is Noun Phrase with process as a head.

f. Research

Word *research* have 694 frequency. First sentence by Participant 5 is “what does research say about assessment...”, it is Noun Phrase with research as head. Next sentence by Participant 12 is “research proposal is a well-organized research plan summary that answers...”, it is noun phrase with research as a head. The last sentence by Participant 12 is “qualitative approach is selected to conduct the research. Creswell stated that...”, it is adjective phrase with researcher as a head.

g. Researcher

Word *researcher* have 414 frequency. Because researcher is a noun, most of the sentence is in noun phrase form. First sentence by Participant 2 is “the instrument of this research is researcher itself...”, it is adverbial phrase with researcher as a head. Next sentence by participant 2 is “contribution for English teacher and further researcher...”, it is noun phrase with research as a head.

h. Scope

Word *scope* have 27 frequency and most of participant use it in form of noun phrase. For example, “scope of the study is the teacher strategies ...” by participant 7, it is noun phrase with scope as a head.

i. Source

Word *source* have 50 frequency. First sentence by participant 11 is “will be used as the main learning source

in learning process...”, it is noun phrase with source as a head. The last sentence by participant 13 is “feedback provided to a learner from any source that contains evidence of learner error...”, it is also noun phrase with source as a head.

j. Technique

Word technique have 62 frequency and most of the participant use it in form of noun phrase. For example, “in qualitative research, the data analysis technique has meaning to organize the result of...” sentence by participant 1, it is noun phrase with technique as the head.

Frequency of using a vocabulary by the writer cannot be separated from the frequency level or sub list, this study focuses on academic vocabulary as the main aspect of the research. There is sub list / level of Academic Vocabulary List,⁵⁹ which means the frequency of the academic vocabulary used by the writer in their work. Smallest the number of sub list, more often the academic vocabulary appear or used by the writer. Vice versa, largest the number of sub list, rarer the academic vocabulary used by the writer. Furthermore, in this sub list by Averil Coxhead, contain all of word families of academic vocabulary, for example *benefit*, so word *benefits* and *beneficial* are member of word list *benefit*. British and American spelling is included in the word families so *analysing* and *analyzing* are both included in the family *analyse*. For example, in sub list 1 of Academic Vocabulary List there are vocabularies such as *benefit*, *approach*, and *context*. The level of sub list of Academic Vocabulary List start from 1 until 10, and for details of the sub list of Academic Word List, see the list in **appendix 3**.

Based on the Table 4.2, there are 10 vocabularies. For the level of the vocabularies according to the sub list of academic vocabulary are explain as follow: word *Research*, *Data*,

⁵⁹ Averil Coxhead, *Sublists of Academic Word List*, Print Book., English Language institute occasional publication 18 ([Wellington, N.Z.] : School of Linguistics and Applied Language Studies, Victoria University of Wellington, 1998), <https://www.victoria.ac.nz/lals/resources/academicwordlist/sublist>.

Researcher, Process, Analysis, Method, Source, and Definition are in sub list 1. For word *Technique* in sub list 3, and for word *Scope* is in sub list 6.

2. Data Analysis on Students' Thesis Proposal

The students remained anonymous throughout the study to guarantee the privacy and confidentiality of the students. The findings of this study are described below, and for details of the students result, see the table in *appendix 2*:

No.	Name	Token%
1.	Participant 1	12.58 %
2.	Participant 2	11.34 %
3.	Participant 3	10.65 %
4.	Participant 4	12.31 %
5.	Participant 5	13.77 %
6.	Participant 6	13.11 %
7.	Participant 7	10.03 %
8.	Participant 8	11.24 %
9.	Participant 9	12.16 %
10.	Participant 10	7.34 %
11.	Participant 11	13.40 %
12.	Participant 12	13.66 %
13.	Participant 13	14.16 %
14.	Participant 14	12.01 %
15.	Participant 15	12.86 %

Table 4.3: Token Percentage of AWL

The data that obtained by researcher after measure all of it from the raw data and interpreted, shows that Participant 1's

proposal document contains 12,58% of academic word list token. The proposal of Participant 2 contains 11.43% of academic word list token. The proposal of Participant 3 contains 10.65% of academic word list token. Next is proposal of Participant 4 contains 12.31% of academic word list token. For the Participant 5's proposal document, it contains 13.77% of academic word list token.

The next is a proposal document of Participant 6, it contains 13.11% of academic word list. The proposal of Participant 7 contains 10.03% of academic word list token. Next is proposal document of Participant 8 contains 11.24% of academic word list token. For the participant 9's proposal document, it contains 12.16% of academic word list token. For the Participant 10's proposal document, it contains 7.34% of academic word list token.

The next is a proposal document of Participant 11, it contains 13.40% of academic word list token. The proposal of Participant 12 contains 13.66% of academic word list token. Next is proposal document of Participant 13 contains 14.16% of academic word list token. For the Participant 14's proposal document, it contains 12.01% of academic word list token. For the Participant 15's proposal document, it contains 12.86% of academic word list token.

From those tables confirm that almost all of the participant obtains high score except for participant number 10 which only obtain 7.34% score. It indicates that almost all of the participants' proposal documents except for participant number 10, can be categorized as Academic Work. Since it has been stated that the percentage of AWL must be approximately 10% which indicate the article or work have fulfilled the criteria of Academic work. This academic word list of Coxhead makes up about 10% of the running words in the text.⁶⁰

⁶⁰ Nation, *Learning Vocabulary in Another Language*, 305.

B. Discussion

In order to have the same interpretation between the reader and the researcher toward the finding of the research, this part discusses those findings by reflecting on some theories related to each following problem. The result of the research was known correctly after doing an analysis of the data. The presented data and the explanation above draw the academic vocabulary use of students' proposal document. Thus, the discussion elaborated based on the findings of research question.

1. Academic Vocabulary Used in Proposal Writing

The research question aims to investigate the vocabulary used by the students in their proposal documents. As explained in finding, to get the data, firstly researcher have to analyse the percentage of academic vocabulary token in each document from each participant. To get the academic vocabulary used by each participant, the researcher needs to classify the academic vocabulary according to their proposal documents. To generalize the data, the researcher takes the academic vocabulary which exist in their document or used by all of the participant by sorting it.

The finding proved that there are 10 academic vocabulary that used in all of the proposal document such as *Research, Data, Researcher, Process Analysis Method, Technique, Source, Definition, and Scope*. Each word has been interpreted and analysed by crosscheck it through each participant's document. This finding brings the assumption that to pay more attention to these AWL words. this statement is supported by some theorist which did a lexical analysis in corpus.⁶¹

2. Data Analysis on Students' Thesis Proposal

As the main objective of the study, analysing academic vocabulary used in students' proposal writing the researcher

⁶¹ Olga Mudraya, "Engineering English: A Lexical Frequency Instructional Model," *English for Specific Purposes* 25, no. 2 (January 2006): 235–256.

aims to correlate the finding in this case to some theories. The finding proved that most of students' token percentage are quite high. This finding brings the assumption that most of the proposal documents are categorized as an academic work. This statement is supported by some theorist which affirm that the document categorized as academic work, since the AWL accounted about 10% of the total tokens.⁶²

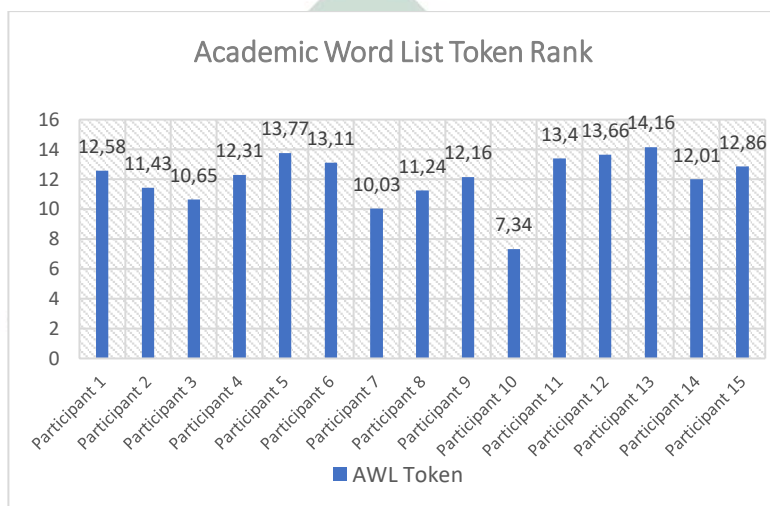


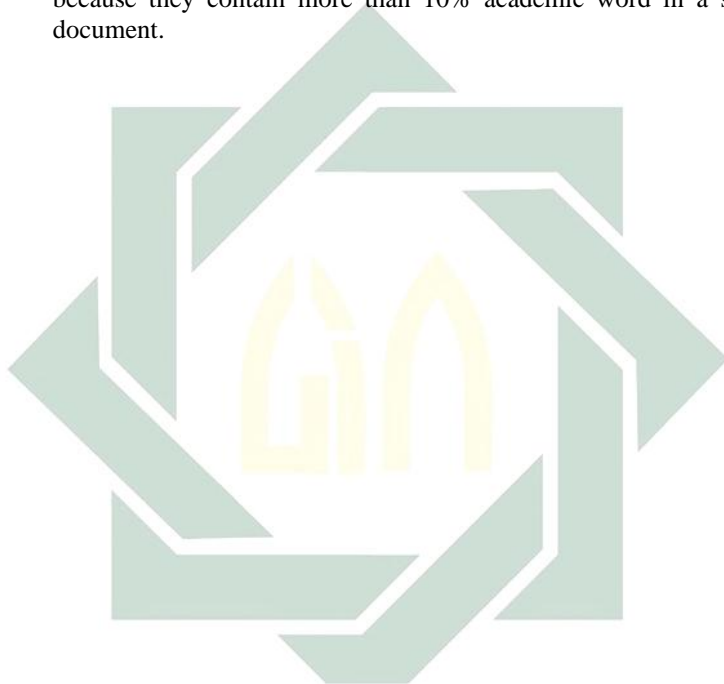
Figure 4.1: Academic Word List rank

From Figure 4.1, it can be observed that almost all of the students have a high score. Start from token score 14.16% obtained by participant number 13, then participant number 5 in second place with token score 13.77%, the last is with token score 13.66% obtained by participant number 12 as the top three out of 15 participants. Meanwhile, participant number 10 who obtain token score 7.34% is the only one participant that cannot accomplish a minimum score of Academic Word List coverage, since the

⁶² Mozaffari and Moini, "Academic Words in Education Research Articles."

Academic Word List approximately have to cover about 10% in a single document.

The findings of this study indicate that academic vocabulary mastering by seventh semester students are good enough. It was proven from 14 out of 15 data categorized into academic article because they contain more than 10% academic word in a single document.



CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

Based on the research finding and discussion, the use of academic vocabulary by seventh semester students are quite high with 11.92% of academic vocabulary token from 66419 of running words. Also, it is proven from 14 out of 15 documents are categorized as an Academic Work / Article. Since academic vocabulary covered approximately 10% in each document. The top three of highest score achieved by Participant 13 with score 14.16% of academic word list token, then Participant 5 with score 13.77% of academic word list token, and the last is Participant 12 with score 13.66% of academic word list token. Furthermore, students of English teacher education department of UIN Sunan Ampel Surabaya are supposed to use academic writing in seventh semester especially in thesis proposal seminar class. They have been suggested to use the academic vocabulary since the beginning of the semester when they are in paragraph writing course considering that it can be a preparation to write the thesis as the final academic task.

Following up the importance of academic vocabulary in academic writing, the researcher comes up with the idea of identifying the academic vocabulary used by seventh semester students of English teacher education department of UIN Sunan Ampel Surabaya based on their thesis proposal.

The items or aspects that can be measured in vocabulary can be through Token, Type, Lemmas, and Word Families, in this study, the researcher use Token as the measured aspect because the researcher need to investigate about how much vocabulary is used in participants' thesis proposal. The result shows that fourteen out of fifteen data that have been measured.

B. Suggestion

After conducting the research about academic vocabulary use in thesis proposal of seventh semester students of UIN Sunan

Ampel Surabaya, the researcher suggest some significant things as follow:

1. For the Lecturer

The researcher suggests that pay more attention on academic vocabulary used, and it should be taught in early semester. Since, the students expected to make a journal or article which is in academic form and to be consumed by public.

2. For the Students

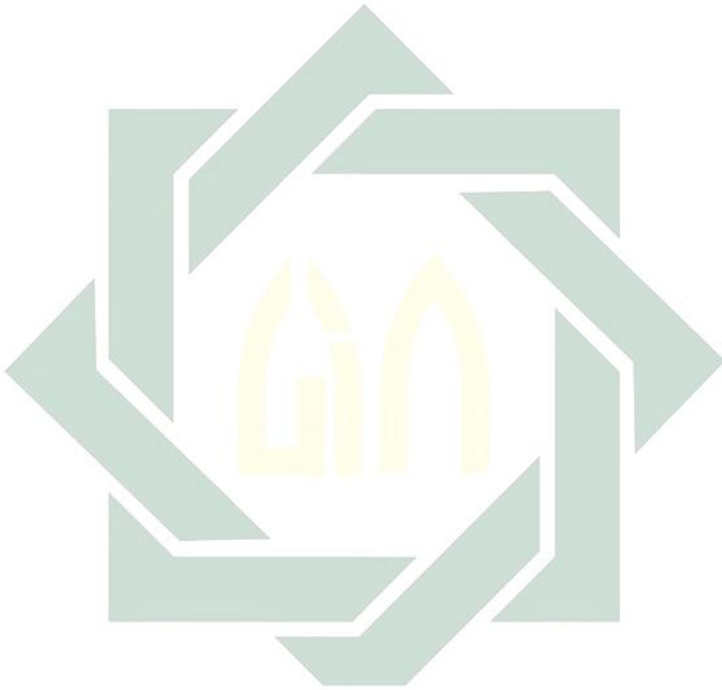
The researcher suggests that students have to learn about academic vocabulary considering that they have weakness of academic vocabulary and try to apply it in term of making some scientific journal or article. Beside that students may also have knowledge about any other information that might help the students to learn the word, or about typical combination of words such as adjectives or verbs typically associated with a noun or nouns, adverbs, and prepositions that associated with a verb.

3. For the Future Researcher

The researcher suggests that future researcher to conduct a research using English Students' thesis as the primary data. However, the future researcher should at least consider that not all students' theses are worth enough to be researched. Some students may be able or have a courage to revise it, but some may not. Therefore, the researcher suggests for the future researcher to conduct a pre-research related to students' plan of continuing their study or not. The research will be significant if the students plan to continue their study considering that they can learn from their weaknesses of academic vocabulary requirement in their previous thesis.

The researcher also suggests to the further researcher to conduct a research using another source of corpus such

as a lexical analysis using BNC (British National Corpus), SEEC (Student Engineering English Corpus). Furthermore, the researcher also suggests to the future researcher to do a corpus analysis about vocabulary used semantically,



REFERENCES

- Albrechtsen, Dorte, Kirsten Haastrup, and Birgit Henriksen. *Vocabulary and Writing in a First and Second Language*. London: Palgrave Macmillan UK, 2008. Accessed August 20, 2019. <http://link.springer.com/10.1057/9780230593404>.
- Anthony, Laurence. "Laurence Anthony's AntWordProfiler." Accessed June 1, 2019. Anthony, L. (2013). AntWordProfiler (Version 1.4.0) Windows. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Baker, Paul. "Sociolinguistics and Corpus Linguistics" (n.d.): 200.
- Baumann, James F., and Michael F. Graves. "What Is Academic Vocabulary?" *Journal of Adolescent & Adult Literacy* 54, no. 1 (2010): 4–12.
- Coombe, Christine. "Assessing Vocabulary in the Language Classroom" (n.d.): 14.
- Coxhead, Averil. "An Introduction to the Academic Word List" (n.d.): 9.
- . "An Introduction to the Academic Word List" (n.d.): 9.
- . *Sublists of Academic Word List*. Print Book. English Language institute occasional publication 18. [Wellington, N.Z.]: School of Linguistics and Applied Language Studies, Victoria University of Wellington, 1998. <https://www.victoria.ac.nz/lals/resources/academicwordlist/sublist>.
- Creswell, John W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 4th ed. Boston: Pearson, 2012.

- . *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 4th ed. Boston: Pearson, 2012.
- Csomay, Eniko, and Alexandra Prades. “Academic Vocabulary in ESL Student Papers: A Corpus-Based Study.” *Journal of English for Academic Purposes* 33 (May 2018): 100–118.
- Curriculum, DataWORKS. “Vocabulary Development Part 1: Types of Vocabulary.” *Dataworks Educational Research*, July 29, 2014. Accessed August 10, 2019. <https://dataworks-ed.com/blog/2014/07/vocabulary-development-part-1-types-of-vocabulary/>.
- Evans, Stephen, and Christopher Green. “Why EAP Is Necessary: A Survey of Hong Kong Tertiary Students.” *Journal of English for Academic Purposes* 6, no. 1 (January 1, 2007): 3–17.
- Haslerud, Vibecke, and Anna-Brita Stenstrom. “The Bergen Corpus of London Teenager Language (COLT).” *Spoken English on computer: Edited by Geoffrey Leech, Greg Myers, Jenny Thomas* (n.d.).
- Huong, Le Pham Hoai. “A Survey Study on Academic Vocabulary Learning Strategies by EFL University Students.” *Journal of Language Teaching and Research* 9, no. 5 (September 1, 2018): 1009.
- Ihalainen, Ossi, Terttu Nevalainen, and Matti Rissanen. “History of Englishes: New Method and Interpretations in Historical Linguistics.” In *The Grammatical Subject in Educated and Dialectal English: Comparing the London-Lund Corpus and the Helsinki Corpus of Modern English Dialects*, 799. Topics in English Linguistics 10. Johansson, Stig; English Computer Corpora: Selected Papers and Research Guide; De Gruyter Mouton: Berlin, Boston, 1992.

- Jamalzadeh, Mehri, and Azize Chalak. "A Corpus-Based Study of Academic Vocabulary in Physiotherapy Research Articles" (n.d.): 15.
- Lauffer, Batia, and Tina Waldman. "Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English: Verb-Noun Collocations in L2 Writing." *Language Learning* 61, no. 2 (June 2011): 647–672.
- Laurence, Anthony. "Laurence Anthony's AntConc." Accessed August 20, 2019. <https://www.laurenceanthony.net/software/antconc/>.
- Malmström, Hans, Diane Pecorari, and Philip Shaw. "Words for What? Contrasting University Students' Receptive and Productive Academic Vocabulary Needs." *English for Specific Purposes* 50 (April 2018): 28–39.
- Mauch, James. *Guide to the Successful Thesis and Dissertation: A Handbook For Students And Faculty, Fifth Edition*, 5th ed. CRC Press, 2003. Accessed August 2, 2019. <https://www.taylorfrancis.com/books/9781482293081>.
- Mozaffari, Ali, and Raouf Moini. "Academic Words in Education Research Articles: A Corpus Study." *Procedia - Social and Behavioral Sciences* 98 (May 2014): 1290–1296.
- Mudraya, Olga. "Engineering English: A Lexical Frequency Instructional Model." *English for Specific Purposes* 25, no. 2 (January 2006): 235–256.
- Nation, I. S. P. *Learning Vocabulary in Another Language*. Cambridge; New York: Cambridge University Press, 2001. Accessed May 23, 2019. <http://dx.doi.org/10.1017/CBO9781139524759>.
- O'Keeffe, Anne, Michael McCarthy, and Ronald Carter. "From Corpus to Classroom: Language Use and Language Teaching" (n.d.): 333.

- . “From Corpus to Classroom: Language Use and Language Teaching” (n.d.): 333.
- Schmitt, Norbert. *Researching Vocabulary: A Vocabulary Research Manual*. Research and practice in applied linguistics. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2010.
- Schuth, Elisabeth, Judith Köhne, and Sabine Weinert. “The Influence of Academic Vocabulary Knowledge on School Performance.” *Learning and Instruction* 49 (June 2017): 157–165.
- Solan, Lawrence M, and Peter M Tieresma. “Author Identification in American Courts.” *Applied Linguistics* 25 (n.d.): 448–465.
- West, Michael. “General Service List of English,” n.d.
- Zhai, Lili. “A Study on Chinese EFL Learners’ Vocabulary Usage in Writing.” *Journal of Language Teaching and Research* 7, no. 4 (July 1, 2016): 752.
- “BUSD Grade Level Academic Vocabulary,” n.d.
- “Collocation - Wikipedia.” Accessed August 12, 2019. <https://en.wikipedia.org/wiki/Collocation>.
- “Sutarsyah, Cucu & Nation, Paul & Kennedy, Graeme. (1994). How Useful Is Eap Vocabulary for Esp? A Corpus Based Case Study. *Relc Journal*. 25. 34-50. 10.1177/003368829402500203.,” n.d.
- “Types of Academic Writing.” *The University of Sydney*. Accessed February 12, 2019. </students/writing/types-of-academic-writing.html>.
- “What Is a Corpus, What Is Corpus Linguistics?” (n.d.): 18.

“WordSmith Tools.” Accessed August 20, 2019.
<https://lexically.net/wordsmith/index.html>.

“Writing Academic English, 4th Edition.Pdf,” n.d.

