DISTRIBUTION OF ACADEMIC WORD LIST IN NOBEL: JOURNAL OF LITERATURE AND LANGUAGE TEACHING

THESIS

BY:

RIZKY YUMNA PRATAMA

REG. NUMBER A73218074

ENGLISH LITERATURE DEPARTMENT

FACULTY OF ADAB AND HUMANITIES

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2023

DECLARATION

I am the undersigned below:

Name   : Rizky Yumna Pratama

NIM   : A73218074

Department   : English Literature

Faculty   : Adab and Humanities

University   : UIN Sunan Ampel

declare that the thesis entitled:

Distribution of Academic Word List in NOBEL: Journal of Literature and Language Teaching

is my own work, and not a plagiarism/fabrication in part or in whole.

If in the future it is proven that this thesis results from plagiarism/fabrication, either in part or whole, then I am willing to accept sanctions for such actions in accordance with the applicable provisions.

Surabaya, 20 JUNE 2023

Rizky Yumna Pratama

Reg. Number. A73218074

v

APPROVAL SHEET

DISTRIBUTION OF ACADEMIC WORD LIST IN NOBEL: JOURNAL OF LITERATURE AND LANGUAGE TEACHING
by
Rizky Yumna Pratama
Reg. Number A73218074

approved to be examined by the board of examiners of English Literature Department, Faculty of Adab and Humanities, UIN Sunan Ampel Surabaya

Surabaya, 20th June 2023

Advisor 1

Suhandoko, M.Pd
NIP. 198905282018011002

Acknowledged by
The Head of the English Literature Department

Endratno Pilih Swasono, M.Pd
NIP. 197106072003121001

EXAMINER SHEET

This is to certify that the *Sarjana* thesis of Rizky Yumna Pratama (Reg. Number A73218074) entitled **Distribution of Academic Word List in NOBEL: Journal of Literature and Language Teaching** has been approved and accepted by the board of examiners for the degree of *Sarjana Sastra (S.S.)*, English Literature Department, Faculty of Adab and Humanities, UIN Sunan Ampel Surabaya.

Surabaya, July 5th 2023

Board of Examiners:

Examiner 1

Suhandoko, M.Pd.
NIP. 198905282018011002

Examiner 2

Murni Fidiyanti, M.A.
NIP. 198305302011012011

Examiner 3

Tristy Kartika Fi'aunillah, M.A.
NIP. 199303182020122018
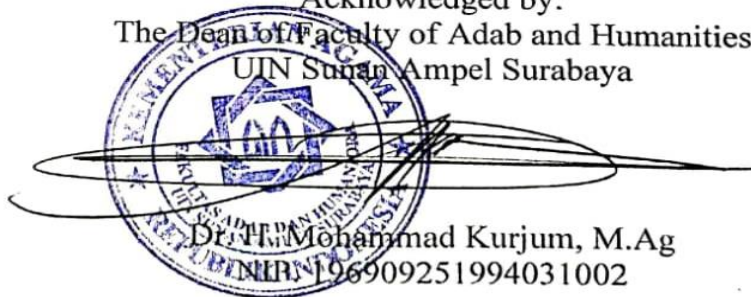
Examiner 4

Endratno Pilih Swasono, M.Pd
NIP. 197106072003121001

Acknowledged by:
The Dean of Faculty of Adab and Humanities
UIN Sunan Ampel Surabaya

Dr. H. Mohammad Kurjum, M.Ag
NIP. 196909251994031002

iv

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Rizky Yumna Pratama

NIM : A73218074

Fakultas/Jurusan : Adab dan Humaniora/Sastra Inggris

E-mail address : pratamarizky201@gmail.com

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

☑ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (..)

yang berjudul :

DISTRIBUTION OF ACADEMIC WORD LIST IN NOBEL: JOURNAL OF LITERATURE

AND LANGUAGE TEACHING

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif 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, 19 Juli 2023

Penulis

(Rizky Yumna Pratama)

ABSTRACT

Pratama, R.Y. (2023). *Distribution of Academic Word List in NOBEL: Journal of Literature and Language Teaching.* English Literature Department, Faculty of Adab and Humanities, UIN Sunan Ampel Surabaya. Advisor: Suhandoko, M.Pd.

The study aims to investigate the distribution of the General Service List (GSL) and the Academic Word List (AWL) in NOBEL: Journal of Literature and Language Teaching. Two problems were analyzed in this study: (1) What is the coverage of academic vocabulary in journal articles published by NOBEL: Journal of Literature and Language Teaching? (2) How does the coverage of academic vocabulary differ in each volume? 78 open-access NOBEL research articles, ranging from volume 7 issue 2 to volume 12, published between 2016 and 2021, were selected and compiled for analysis in this study. The AntWordProfiler 2.0.1 was used to count the frequency and coverage percentage of words from the AWL and GSL word lists.

The total coverage of running words (tokens) in level 1 and level 2 of the GSL accounted for 78.52% of the corpus, while the AWL covers 9.17% of the corpus. Additionally, the researcher presents the top 10 AWL word families that represent the entire articles, showing the word families that were used most frequently in the entire corpus. Furthermore, the analysis of the development of academic coverage from volume to volume is presented, along with the top 10 AWL word families in each volume, which are also provided in this research. The findings regarding the distribution and coverage of academic words in the articles significantly contribute to recognizing the quality of academic word coverage comprehension in the articles.

Keywords: academic word list, general service list, word list coverage, corpus linguistics

ABSTRAK

Pratama, R. 2023. *Distribusi Daftar Kata Akademik di NOBEL Jurnal Sastra dan Pengajaran Bahasa*. Program Studi Sastra Inggris, Fakultas Adab dan Humaniora, UIN Sunan Ampel Surabaya. Pembimbing: Suhandoko, M.Pd.

Penelitian ini bertujuan untuk menginvestigasi distribusi *General Service List* (GSL) dan *Academic Word List* (AWL) di *NOBEL: Journal of Literature and Language Teaching*. Terdapat dua masalah dalam penelitian ini yang akan dianalisis, yaitu: (1) Bagaimana cakupan kosakata akademik dalam artikel jurnal yang diterbitkan oleh *NOBEL: Journal of Literature and Language Teaching?* (2) Bagaimana perbedaan cakupan kosakata akademik di setiap volume? Tujuh puluh delapan artikel penelitian *NOBEL* akses terbuka, mulai dari volume 7 edisi 2 hingga volume 12 yang diterbitkan pada tahun 2016 hingga 2021, dipilih dan dikompilasi untuk dianalisis dalam penelitian ini. Dalam penelitian ini, The *AntWordProfiler 2.0.1* digunakan untuk menghitung frekuensi dan persentase cakupan kata dari dua daftar kata, yaitu AWL dan GSL.

Cakupan total *running words* (token) pada level 1 dan level 2 GSL sebesar 78,52% dari korpus. Sementara itu, AWL mencakup 9,17% dari korpus. Selain itu, peneliti menyajikan 10 famili kata AWL teratas yang mewakili keseluruhan artikel, menunjukkan famili kata yang paling sering digunakan dalam korpus keseluruhan. Selanjutnya, analisis hasil perkembangan cakupan akademik dari volume ke volume disajikan bersama dengan 10 famili kata AWL teratas di setiap volume yang juga disediakan dalam penelitian ini. Temuan mengenai distribusi dan cakupan kata-kata akademik dalam artikel secara signifikan berkontribusi untuk mengenali kualitas pemahaman cakupan kata-kata akademik dalam artikel tersebut.

Kata Kunci: daftar kata akademik, daftar servis umum, cakupan daftar kata, korpus linguistik.

TABLE OF CONTENTS

CHAPTER I INTRODUCTION

CHAPTER II REVIEW OF RELATED LITERATURE

CHAPTER III RESEARCH METHOD

x

LIST OF TABLES

xi

LIST OF FIGURES

xii

This chapter presents the introduction of the research, which discusses the background of the study and objectives of the study, significance of study, scope and limitation of study, and the definition of key terms.

1.1 Background of the Study

Languages have been utilized by humans to interact with each other for thousands of years. Chomsky (2000) stated that a language comprises finite or infinite sentences constructed using a limited set of elements, and native speakers possess the inherent ability to understand and create grammatically correct sentences. From previous generations to the present, languages have developed with a massive number of vocabulary words. Nowadays, English is recognized as an international language used by the majority of people worldwide, whether as a first, second, or foreign language. Throughout history, the usage of vocabulary has also evolved within decades, enabling people to interact with each other. However, the usage of vocabulary in languages differs depending on the context in which they are utilized. This becomes an issue when people are not aware of specific vocabulary words, such as technical, medical, and academic terms.

In the academic world, the application of academic vocabulary in research and study is considered essential for scholars. Nonetheless, some researchers and students face difficulties in comprehending academic vocabulary. Mabruroh (2020) reported in her research focuses on the difficulties faced by Indonesian students studying English, specifically senior students in the English Department

1

during the 2015/2016 academic year. The study reveals that these students encounter challenges when reading academic texts, including problems with vocabulary, understanding the content, grasping concepts and topics, and lacking motivation from their teachers. Surprisingly this study found that comprehension was the biggest difficulty for students in academic reading. Furthermore, the research indicates that these academic reading difficulties not only impact students' comprehension abilities but also diminish their interest in reading.

Hartono & Badara Prima (2021) presented in their research, a total of 168 first-year university students from a private institution that used English-medium instruction (EMI) were included as participants. To assess their vocabulary knowledge, an updated vocabulary level test was administered, while their reading comprehension level was measured using the IELTS academic reading test. The results showed that only 34 out of the 168 participants demonstrated an adequate vocabulary size. This indicates that the majority of participants lacked sufficient vocabulary knowledge and had a low level of proficiency in reading. These deficiencies in vocabulary and reading proficiency could potentially hinder their academic progress and achievement.

Shaw (1991) asserted that the usage of vocabulary becomes a dilemma for native-speaker and nonnative-speaker learners in academic reading and writing. Particularly in Indonesia, students have been taught English for at least 12 years or more, starting from elementary school to university. Nevertheless, the insufficient usage of academic vocabulary in writing academic papers becomes a significant obstacle for both students and researchers. Evan and Green (2007) conducted a survey with undergraduate students at Hong Kong's largest English-

medium university, surveying almost 5000 undergraduates through a questionnaire. The survey results showed that students face obstacles in academic writing and speaking. Several studies have revealed a correlation between vocabulary proficiency and reading comprehension. When students have insufficient vocabulary, they tend to struggle to comprehend the content they read, which hinders their understanding of the context (Nation & Waring, 1997).

Numerous researchers have conducted studies on the vocabulary size of native English speakers and second language (L2) learners. Nation and Waring (1997) indicate in their research that native English speakers typically recognize approximately 20,000-word families and expand their vocabulary comprehension by around 1,000-word families annually on average. Several studies on vocabulary acquisition (Goulden, Nation, & Read, 1990; Hirsh & Nation, 1992) reported that university students also acquire a vocabulary size of around 20,000-word families, excluding proper names, compound words, abbreviations, and foreign words. Additionally, Groot (2000) stated that second language learners require a vast vocabulary size to achieve reading capacity.

The recognition of the importance of vocabulary knowledge in utilizing language in different situations has been a constant focus for vocabulary researchers in linguistics. Coxhead and Nation (2001) have generally classified English vocabulary into four categories: general service words, academic words, technical words, and low-frequency words. It is highly suggested for researchers and university students to utilize academic vocabulary in their written research rather than relying on general service words. In this study, the researcher will analyze the coverage of academic words based on corpus-based analysis.

Damico and Ball (2019) stated that corpus linguistics is a subfield of linguistics that involves the deliberate collection of naturally occurring samples of spoken or written language for linguistic analysis. Corpus linguistics has been described as a methodology rather than a theory for conducting linguistics. It involves the creation and production of natural language samples for linguistic analysis from various theoretical perspectives. The analytical methods used to describe corpora rely on the researcher's questions and can be associated with a variety of linguistic theories, ranging from descriptive grammars to sociolinguistic approaches. Numerous researchers have developed corpora based on their specific purposes, such as academic, technical, and general corpora.

Several corpus linguistics researchers have developed word lists to identify the coverage of corpora derived from various sources, such as papers, journals, books, and articles. West (1953) developed the General Service List (GSL) from a corpus of 5 million words with ESL/EFL learners in mind. This list comprises the 2,000 most useful word families in English, selected based on criteria including frequency, ease of learning, coverage of valuable concepts, and stylistic level. Coxhead (2000) also developed an academic corpus known as the Academic Word List (AWL) for academic purposes. Her research involved the collection of word lists from 414 academic texts by more than 400 authors, with approximately 11,666 pages of text comprising 3,513,330 tokens (running words) and 70,377 types (individual words). The main corpus was subdivided into arts, commerce, law, and science. In the current study, the focus is on the coverage of AWL in NOBEL journal articles, with the GSL coverage provided as additional data.

Previous studies have utilized AWL for analysis. Hyland and Tse (2007) conducted research on an academic corpus consisting of research articles, academic books, master's theses, doctoral dissertations, and scientific letters. The study found that out of a total of 3,213,477 running words, 340,035 AWL items were identified, resulting in an average coverage of 10.6%. The AWL coverage percentage varied among disciplines, with Engineering at 73.3%, Social Sciences at 77.0%, and Sciences at 69.0%. Additionally, the GSL coverage was relatively higher, with an overall average of 84.7% of the corpus.

Another research study focused on examining academic words. Khani and Tazik (2013) collected 1,553,450 running words from 240 Applied Linguistics research articles with the aim of composing an academic word list for the subject of research. The corpus consisted of 32,479 words types and 2,409 words families. Of the running words, 1,165,088 were among the first and second 1,000 words of the GSL 2000, accounting for 76.4% of the total corpus. Additionally, 194,355 running words were in Coxhead's (2000) AWL, accounting for 11.96% of the total tokens. This finding showed a higher AWL coverage compared to the approximate 10% coverage suggested by AWL.

Another study focused on word list dispersion. Chanasattru and Tangkiengsirisin (2017) analyzed English social science research articles from 11 journals obtained from the ScienceDirect website. The analysis utilized the New General Service English (NGSL) and AWL. The total coverage of NGSL accounted for 73.20% with 303,425 running words, while the coverage of AWL accounted for 13.86% with 53,675 running words. This analysis indicated a higher coverage of AWL, approximately 10% of the corpus. Based on the

aforementioned studies, research investigating the distribution of academic words in academic works is relevant to the current study. However, there is limited research on investigating the distribution of academic words in journal articles, particularly in Indonesian journal publishers. Therefore, this study focuses on analyzing the academic word coverage and listing the most frequently occurring academic words in NOBEL journal articles, specifically volume 7 issue 2 to Volume 12, published between 2016 and 2021.

The reason for choosing NOBEL as the data source is that no previous studies have conducted research on investigating the distribution of academic words in NOBEL through corpus-based analysis. Volume 7 issue 2 was chosen because this issue has been started to be accredited by Sinta, a digital academic data repository that provides quick and simple access for reviewing the performance of Indonesian researchers, institutions, and journals. It also assists institutions in analyzing research standards and recognizing their research strengths, fostering partnerships and cooperation. Additionally, Sinta aids in examining study trends and provides directories of field-specific experts. Journal articles from NOBEL are suitable data for this research because they are related to academic vocabulary and academic researchers.

1.2 Problems of the Study

In line with the background above, the problems of study are formulated as follows:

1. How is the coverage of academic vocabulary in journal articles published in NOBEL?

2. How is the coverage of academic vocabulary dispersed in each volume?

1.3 Objectives of the Study

In line with the problems of the study above, this study aims to

1. Identify the distribution of academic words dispersing in entire articles through AWL as the analysis-based

2. Identify the distribution of the academic words dispersing in every volume

1.4 Significances of the Study

This research hopefully provides a contribution to the field of linguistics, particularly in the corpus-based analysis of the academic corpus. Considering that the application of the academic vocabulary is essential for academic researchers. This research is proposed to identify and explain the finding of distribution of academic words in the NOBEL journal articles. The result may help for the future researcher understanding the coverage of academic words in NOBEL journal articles, and also how to analyze the data and present the analysis of the finding. this research hopefully provides data of the analyzed articles in NOBEL and for the editor of NOBEL journal hopefully perceives the distribution of academic words across the volumes by knowing the most occurred academic words across all the selected volume and the development of the academic word coverage in each volume from year to year.

1.5 Scope and Delimitations

The scope of this study focuses mainly on analysis of the distribution of academic words in NOBEL journal articles while general English words as additional information will be provided in the analysis. Prasetya et al. (2020)

found that corpus linguistics employs computer technology to gather language data, systematically collects speech samples, uses specific procedures to process language datasets, and examines significant collections of written and spoken texts to represent various language domains within applied linguistics. While in this research using the written based text as the basis data which is the journal articles compiled as corpus research data. The researcher presents the overall coverage calculation of the analysis, frequent words in entire articles, sub-list of academic word, presentative data from articles, and most frequent word in every volume. The delimitation of the study only using the AWL and GSL list as the analysis-based. For calculating the data using the computer software called AntWordProfiler made by Laurence Anthony to help identifying the data. The researcher selects only NOBEL journal articles volume 7 issue 2 (2016) – volume 12 (2021). The NOBEL published firstly journal articles since 2011 but the researcher selects the data from 2016 because the volume was started to be indexed by SINTA accreditation.

1.6 Definition of Key Terms

In order to clarify the key terms used in this study, some definitions are put forward:

AWL is Academic Word List. Collection of academic word developed by Averil Coxhead.

GSL is abbreviation of General Service List. Collection of general English word developed by West.

NOBEL is Journal of Literature and Language Teaching published by English Literature Department UIN Sunan Ampel.

AntWordProfiler is computer software tool analysis to calculate the distribution of word list in corpus.

Word list is a reference list utilized to analyze the data in AntWordProfiler software

Sub-list is a categorized list created by Coxhead to classify the most occurred words in her AWL list.

Frequency is a word occurs mostly in the corpus after calculated by the software

Token is also called running word. The total number of words calculatingly detected by the coverage of word list in AntWordProfiler.

Group is collection of headwords of AWL and GSL which detected by the software analysis.

Type is the member of word families detected in the result of calculation of the analysis.

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter presents the theories relevant to the current research topic. The researcher provides several related theories which include Corpus Linguistics, Word List, Academic Word List, Applied Linguistics Academic Word List, AntWordProfiler.

2.1 Corpus Linguistics

The method of research known as corpus linguistics emphasizes the analysis and study of language using massive archives of texts written in natural language, or corpora. Corpora have carefully gathered collections of written or spoken materials that are organized and sampled from a variety of sources and genres. (Prasetya et al., 2020) concluded from previous researches that corpus linguistics applies a variety of techniques to study language. First, it utilizes modern computer technology to gather language data. It also includes the systematic collection of various speech samples. Additionally, particular procedures and techniques are used in corpus linguistics to process and create language collections of data. In conclusion but not least, it investigates significant collections of written and spoken texts that are specifically chosen to represent different language domains, including casual conversations or academic writing, as a subfield of applied linguistics.

2.2 Word List

Word lists categorized by the frequency of occurrence within a specific text corpus are valuable tools for vocabulary acquisition. These lists, whether

arranged by levels or as ranked inventories, aid in vocabulary learning by providing a systematic approach (Nation, 1997). However, it is important to note that frequency lists are primarily intended for course developers rather than learners themselves. They serve as a reference to ensure that common words are appropriately addressed and not overlooked in language instruction. Several challenges exist when constructing frequency lists, including considerations related to the corpus content, register, and the definition of what constitutes a "word." While manual word counting has a long history, the advent of natural language electronic processing and the analysis of vast corpora, such as the SUBTLEX megastudy utilizing movie subtitles, has significantly advanced this field of research.

In computational linguistics, a frequency list is a sorted compilation of word types along with their frequencies, typically based on the number of occurrences within a specific corpus. This information allows for the derivation of word ranks within the list (Nation, 1997). Applying concordance software, which comprehensively examines every word in the corpus, a frequency list of words can be generated from a corpus. The overall number of occurrences (tokens) and the number of different word kinds are calculated using this search method. Based on this research, the program creates a frequency list that may be displayed either in alphabetical order or in descending order of frequency (Evison, 2010). This indicates that word lists perform an essential function in linguistic studies that attempt to provide information on the frequency with which words occur in particular areas. The generated word frequency lists can aid researchers in

assessing regardless of these words are relevant to their research or should be disregarded.

2.3 Academic Word List

The current study applying the method of analysis based on academic word list. Coxhead (2000) developed the Academic Word List (AWL) was compiled from a corpus of 3.5 million running words of written academic text. It was developed by analyzing the range and frequency of words that occur beyond the first 2,000 of the GSL developed by West (1953) most frequently used words in English. The AWL consists of 570-word families, which account for approximately 10.0% of the total words in academic texts. In comparison, the AWL represents only 1.4% of the total words in a similarly sized collection of fiction. This discrepancy in coverage indicates that the list primarily comprises academic vocabulary. By identifying the words encountered by university students in a variety of academic texts, the AWL assists learners in prioritizing the words most important for academic study.

Table 2.1 Example of The AWL Sub-list

No	AWL Sub-list 1	AWL Sub-list 2	AWL Sub-list 3	AWL Sub-list 4
1	Analyse	Achieve	Alternative	Access
2	Approach	Administration	Circumstance	Attribute
3	Area	Acquire	Comment	Debate
4	Assess	Appropriate	Compensate	Dimension
5	Assume	Assist	Component	Implement
6	Authority	Conclude	Considerable	Impose
7	Available	Conduct	Constant	Implicate
8	Concept	Community	Constrain	Integrate
9	Consist	Compute	Contribute	Retain
10	Constitute	Credit	Deduce	Investigate

Additionally, the table above are the samples of AWL sub-list. She also classified the word into sub-list, sub-list is collection of academic word and its families. Sub-list 1 consist of the most common words in the AWL. Sub-list 2 consist of the next most common words and so on. The complete collection of sub-lists can be accessed in

https://www.wgtn.ac.nz/lals/resources/academicwordlist/sublist. the list serves as a valuable foundation for further research on academic vocabulary. The corpus used in the study comprised 414 academic texts written by over 400 authors, containing 3,513,330 tokens (running words) and 70,377 types (individual words), spread across approximately 11,666 pages. It was divided into four sub corpora,

namely arts, commerce, law, and science, each consisting of around 875,000

running words and further subdivided into seven subject areas.

2.4 General Service List

There are only few references of the General Service GSL list. The

General Service List (GSL) was developed by West in 1953 and consists of 2,000

words that are considered highly useful for English language learners. The

selection process considered factors such as frequency, although the list does not

consist of the most common 2,000 words. Each word in the GSL represents a

word family and is loosely defined by West. Frequency numbers were derived

from Thorndike and Lorge (1944) and are provided for different word meanings.

The GSL has had a significant influence on educational materials, including

graded readers. Although the GSL itself is no longer in print, texts based on it are

still available. For a more detailed discussion on the GSL and word lists in

general, refer to Nation (1990, pp. 21-24) and Carter and McCarthy (1988,

Chapter 1). The following data are the example of the first 1000 words and the

second 1000 words, the words were selected randomly with the headword as

presented below:

Table 2.2 The First 1000 Words of the GSL

Able	Break	Bridge	Bread	Could
About	Carry	Class	Clean	Detail
Attempt	Command	Difficult	Except	Excellent
Average	Fix	Foreign	Form	Fortune
Dream	Factory	Gather	Fashion	Good
Ground	Influence	Interest	Introduce	Happy
Join	Just	Keep	Kind	Lack
Lady	Manner	Many	Never	Neck

Table 2.3 The Second 1000 words of the GSL

Abroad	Accident	Accustom	Bake	Balance	Basket
Coal	Clay	Clever	Defend	Delicate	Depend
Empty	Encourage	Especial	Fade	Fate	Fence
Garage	Generous	Grain	Hammer	Harvest	Heap
Insect	Inside	Interfere	Joy	Joke	Jump
Kick	Knot	Knee	Loaf	Lock	Luck
Manage	Meanwhile	Melt	Nice	Neglect	Nose
Obey	Onto	Ocean	Pad	Passage	Paste

The tables example can be useful for language learners, educators, and researchers who aim to enhance their vocabulary or analyze word frequency in the English language. Both tables consist of various common words used in the English language. The General Service List (GSL) is a valuable resource for English

language learners, educators, and researchers as it provides a comprehensive list of commonly used words in the English language. The complete data can be accessed in https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists/general-service-list/general-service-list-headwords.pdf

2.5 Applied Linguistics Academic Word List

Previous researchers have conducted at investigating the academic words in applied linguistics researches. Matinparsa et al. (2022) conduct research in finding the distribution of academic words in applied linguistics articles. they developed a corpus of research articles was compiled from 20 prominent journals in applied linguistics. These journals were carefully selected based on their significance in the field and were published between 2010 and 2020. The corpus consisted of 7,383 research articles, providing a substantial amount of data for analysis. The articles covered various aspects of applied linguistics, including second/foreign language learning/teaching and practical application of language theories. By converting the articles into text files, the corpus was processed using specialized software, resulting in a total word count of 48,279,142 running words. the researchers utilized the AWL as the corpus-based analysis to identify the most occurred words in the research articles.

The research presented the top 50 most frequent words in the corpus, consisting of Academic Word List (AWL) words and non-GSL/AWL words. The frequency shows the number of occurrences of each word in the corpus. The most frequent word is "research," with a frequency of 160,284 followed by "participants" and "task." These words are commonly used in academic writing

and reflect the focus of the research articles in the corpus. Overall, these top 50 words account for approximately 3.77% of the total words in the corpus, indicating their significance and prevalence in the analyzed academic texts. The researchers reported the outcomes of the appearances of applied linguistics vocabulary obtained from the research data, which consisted of applied linguistics research articles from internationally recognized journals. This compilation of academic words in applied linguistics may be useful for other researchers in determining the scope and frequency of word occurrences. The applied linguistics academic word list can be accessed here https://assets.researchsquare.com/files/rs-2092705/v1/27cf29ec755d70c953b6abfe.docx

2.6 AntWordProfiler

In this research project utilize AntWordProfiler for analyzing. (Anthony, 2022) a software called AntWordProfiler was invented to analyze text data articles and determine their coverage. Comparing the text corpus with predefined reference lists of vocabulary was the method of analysis utilized by the software. Its purpose was to examine the appearance or disappearance of particular words from the word lists and to calculate various metrics, including percentage, frequency, range, and coverage of the reference word lists within the corpora. In addition, the software provided statistical information, including the total number of words (tokens), member of word families (types), and headwords detected in the corpus.

Figure 2.1 AntWordProfiler Software
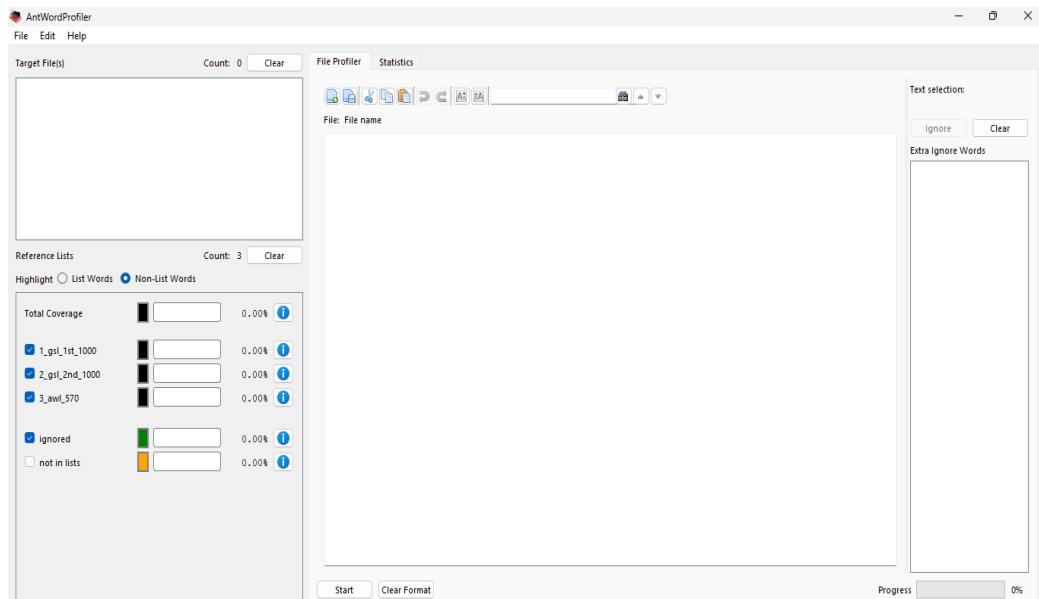
CHAPTER III

RESEARCH METHOD

This chapter presents about: research design, data collection, research data, data source, instrument, data collection technique, and data analysis technique.

3.1 Research Design

The researcher applied corpus-based word list measurements to examine the academic words identified in the corpus of English-language NOBEL Journal articles. The examination employed AntWordProfiler to quantify the corpus texts with the reference word list employed in the tools to determine the frequency of word occurrences. The data is conveyed in a format of tables together with examples and explanations generated directly from the corpus article text.

3.2 Data Collection

The researcher applied the journal articles as the main data to be analyzed for the current study. The following instruction is explained below:

3.2.1   Research Data

The research data are articles from NOBEL: Journal of Literature and Language Teaching, which consists of 78 articles from Volume 7, Issue 2 (2016) to Volume 12 (2021), as the data source. This scholarly journal is published twice a year by the English Department, Faculty of Adab and Humanities, Universitas Islam Negeri Sunan Ampel Surabaya. The journal is abstracted and indexed by reputable platforms such as Google Scholar, Garuda, DOAJ, Moraref, and Indonesia One Search (IOS). Additionally, this publication has been verified by the National Journal Accreditation (ARJUNA).

Table 3.1 NOBEL Journal Articles

No	Volume	Number of articles
1	Volume 7 issue 2	6 articles
2	Volume 8	12 articles
3	Volume 9	12 articles
4	Volume 10	12 articles
5	Volume 11	18 articles
6	Volume 12	18 articles
Total	6 Volumes	78 articles

3.2.2 Data Source

The researcher selectively chose the articles volume 7 issue 2 (2016) –
volume 12 (2021) as the source of the research data which derived from NOBEL:
Journal of Literature and Language Teaching is retrieved from
https://jurnalfahum.uinsby.ac.id/index.php/nobel

3.2.3 Instrument

The researcher applied a specific tool for analyzing data in order to gather
data. AntWordProfiler, software developed by Anthony (2022) and established
publicly accessible for free, was applied in the study to evaluate the vocabulary
level and level of complexity in articles. In this investigation, the software was
used to evaluate and quantify the data.

3.2.4 Data Collection Technique

The researcher collected the articles by downloading online the data which
selected open-access journal articles volume 7 issue 2 – volume 12 comprised 78
articles from the NOBEL : Journal of Literature and Language Teaching website.
These selected journal articles were all written in English containing literature,
linguistics, and language teaching content. After completely downloaded the
articles, the researcher stored and classified the file of the journal articles based on

their volume in computer folder. This study following the research they conducted (Chen & Ghe, 2007) demonstrated standardization was implemented to eliminate uncountable parts, such as charts, diagrams, bibliographies, equations, text headers, footnotes, the writer's name, associates, and several other features of the articles that concordance is unable to process. The data retrieved from the Journal were in pdf forms. The standardized PDF article files required to be converted to UTF-8 plain text format to minimize conversion issues with non-English words. After the text files were error-checked, the data were processed in two corpora, the first corpus consisting the emerged of entire articles of journal, while the second corpus was created a sub-corpus of each volume.

3.3 Data Analysis Technique

In this research, the researcher utilized the computer software called AntWordProfiler by Anthony Lawrence to examine the documents including AWL as the main reference word list and GSL as additional reference word list to calculate the distribution of academic and general vocabularies in documents. In this case, the researcher inserts all the data into the AntWordProfiler software and identify them as explained below:

1. For responding to the first research problem, the researcher opened the AntWordProfiler software to calculate the data

2. The researcher inserted the reference list which means the AWL and the GSL list.

3. The researcher inserted the plain text data corpus of entire volume into Target File box in the software.

4. The researcher activated the Academic Word List by Coxhead in reference list box. The AWL was used as the standardization of examining academic words to analyze the data. The researcher also activated to use general service list to identify common English words.

5. The data processed to be analyzed and the outcomes resulting in the form of statistic data, and the occurrence of word list.

6. The result of calculation in statistic data comprised of tokens, percentage of tokens, types, percentage of types, Headwords, list of the most frequent academic words.

7. The researcher obtained the data from the output of the software that processed in Microsoft Excel form.

8. The results of statistics were converted into data table in this research

9. The resulting frequent words were saved in Microsoft Excel, the frequent words were still scattered. The researcher organized the words by eliminating the word types of headwords, but the accounted word types were accumulated into the headword, not only the headwords were only presented in the data table, but also they consist of their word families.

10. Then the researcher classified the most frequent words based on the highest number of most occurred words existing in the entire corpus.

Furthermore, for responding to the second research problem, the researcher conducted the same data analysis technique as the instruction above. But the difference is only focus to calculate each volume of the total 6 volumes.

3.3.1 Discussing the data

The researcher obtained the analyzed data from the software. In order to interpret the data the researcher imported the statistic data only selected the data in statistic such as token, token percentage, cumulative token percentage, type and group from the software as depicted below:

Figure 3.1 Coverage Statistic

From the picture above, the researcher edited and modified the statistic data of all volumes that will be described in chapter 4.

The researcher collected only the word frequency of AWL, from the outcome of the software analysis still in raw result. So the researcher classified the word based on their headwords which contain their families in excel form. Here are the sample of raw result of word frequency:

Figure 3.2 The Word Frequency of AWL

After classifying the word frequency of AWL in all volume based on their

headwords and their highest number of occurrences. The data will be presented

and explained in Finding of the first problem of the study in chapter 4. This

method also implemented for the second research question which is the number of

word occurrence in every volume. The word occurrence in every volume will be

presented and explained in finding of the second problem of the study in chapter

4. The researcher also presented the data table of the pattern of the statistic AWL

and GSL coverage in every volume that will be presented in finding of the second

problem of the study in chapter 4.

CHAPTER IV

FINDINGS AND DISCUSSION

This chapter presents the result of the study that were conducted according to the method explained in chapter 3. The first section presents the findings of the study about the coverage of academic vocabulary used in Journal Articles. The second section presents the discussion of the findings on analysis of academic vocabulary coverage.

4.1 Findings

This section reports the findings from the data analysis conducted in response to the research question, "What is the extent of academic vocabulary coverage in the journal articles published by NOBEL: Journal of Literature and Language Teaching?" The findings focus on two aspects: first, the use of academic vocabulary in the selected journal articles from Volume 7, Issue 2 to Volume 12, and second, the data analysis procedure. The corpus is comprised of 78 documents extracted from the preceding volumes and compiled into a singular dataset. Using the software application AntWordProfiler, the data was analyzed, and the results are described below.

4.1.1 The Coverage of AWL in NOBEL Journal Articles

In this section, the researcher responds to the first research problem of this study. The researcher conducted the analysis of the distribution of academic words in all volumes based on AWL as the tool analysis. The following is the explained data:

Table 4.1 The Statistic Coverage in All Volume

LIST	TOKEN	TOKEN%	CUMTOKEN%	TYPE	GROUP
GSL 1	253,303	73.95%	73.95%	3,048	978
GSL 2	15,662	4.57%	78.52%	1,725	785
AWL	31,398	9.17%	87.69%	1,880	551
NON GSL/AWL	42,179	12.31%	100%	9,267	9,267
TOTAL	342,542	87.69%		15,920	11,581

The table above displays the lexical profile of the journal articles in NOBEL

based on the General Service List (GSL) and the Academic Word List (AWL).

The first 1000 most frequent words in English from the GSL accounted for

253,303 tokens, representing 73.95% of the corpus. The second 1000 most

frequent words from the GSL accounted for 15,662 tokens, comprising 4.57% of

the corpus. Overall, the GSL lists covered a total of 78.52% of the corpus.

Additionally, the AWL accounted for 31,398 tokens, which corresponded to

9.17% of the entire corpus.

The number of types, which indicates the total count of word families,

varied across the word lists. The first GSL list consisted of 3,048 word types, the

second GSL list had 1,725 word types, and the AWL list contained 1,880 word

types. In total, there were 15,920 types, including both accumulated types and

those not found in the GSL or AWL.

The groups refer to the total count of headwords in the data. The first GSL

had 978 headwords, the second GSL had 785 headwords, and the AWL had 551

headwords. Including both accumulated groups and those not found in the GSL or

AWL, there were a total of 11,581 headwords.

The cumulative coverage of the GSL and AWL accounted for 87.69% of the corpus, indicating that 12.31% or 42,179 tokens were not included in these word lists.

Table 4.2 The AWL frequency in all Volumes

No	Item	Frequency	Sub-list
1	Research	1208	1
2	Analyse	770	1
3	Text	655	2
4	Identify	616	1
5	Strategy	568	2
6	Process	565	1
7	Culture	524	2
8	Data	507	1
9	Function	420	1
10	Media	391	7

The table above is result of the AntWordProfiler calculation as in explanation below: the researcher only selects arbitrarily the most top 10 words frequencies occur in all volumes. The frequency is how many times word appear. Meanwhile the sub-lists are the most frequently occurring member of word families in the academic word list by Coxhead. The academic word list has been separated into sub-lists based on the frequency of occurrence of the words. Sub-list 1 words appear more frequently in the corpus than the other words in the list. Sub-list 2 and their following appears with their lower of word frequency. In this research, the researcher review how is the academic vocabulary used in all volume journal articles. Here are the details found as following:

1. **"Research"**

   In the study, the word "research" and its families were found to occur 1.208 times in the data. The words associated with "research", according to Oxford Dictionary are (1) *research* as singular noun, (2) *researches* as plural noun, (3) *researcher* as -er suffixed singular noun, (4) *researchers* as -er suffixed plural form, (5) *research* as base verb, (6) *researched* as past or past participle form of verb, (7) *researching* as present participle form of verb, (8) *researching* as a gerund. In the study, the researcher found all words associated with *research* taking the forms such as *research, researched, researcher, researchers, researches, researching*. The following is the explanation of data sample.

   Data set 1

   - This ***research*** relates to cultural studies since the literary work is popular literature." (Volume 7 issue 2)
   - "The researcher conducts this ***research*** since the story is interesting to read." (Volume 8)
   - "For example of omission errors that were found in the ***research*** data." (Volume 9)

   In the data set 1, we can see how "*research*" occurs in the articles taking the forms of singular noun either as subject, object, or noun phrase in the sentence. The word *"research"* is used in different ways in the provided statements. It refers to a study exploring the connection between a literary work and popular culture in cultural studies. It also represents an individual's investigation driven by their personal interest in a story. Furthermore, it describes the data analyzed, specifically highlighting the identification of omission errors. Overall, *"research"* encompasses scholarly investigations, personal explorations, and data analysis within a research context.

Data set 2

- "The *researcher* uses the theory from Richards (1974)." (Volume 9)
- "According to Prior (2008), interpreting qualitative method is a method in which the *researcher* construes meaning from research findings and use the *researcher's* assumption/interpretation to provide a context to help the readers understand the result of the study." (Volume 10)
- "These people could help the *researcher* to understand the data better." (Volume 11)

In the data set 2, we can see how "*researcher*" occurs in the articles taking the forms of -er suffix singular noun either as subject, object or noun phrase. The word *"researcher"* in the statements refers to an individual actively engaged in conducting research, utilizing theories, interpreting qualitative methods, and seeking assistance to understand the data better. They play a central role in carrying out the research process and analyzing the results.

Data set 3

- "There has been a substantial upsurge in *researching* the effectiveness of MALL in the ESL/EFL listening classroom." (Volume 11)
- "Tomlin (1986) introduced an attentive-based proportional study, *researching* 402 languages." (Volume 10)
- "Numerous *researchers* have been interested in doing study about disagreement such as studies in academic setting and studies in outside of academic setting." (Volume 8)
- "Considering SFLT as a new perspective in grammar, it has attracted many *researchers* to conduct the research in this field." (Volume 9)

In the data set 3, we can see how "*researching*" occurs in the articles taking the forms of gerund in the sentence. While *"researchers"* taking the forms of plural noun either as subject or object. In summary, *"researchers"* refers to individuals who conduct research, while *"researching"* represents the active process of conducting research itself. Both terms highlight the efforts made by scholars and

experts to contribute to the advancement of knowledge and understanding within their respective fields.

Data set 4

- "This is a library-based study since the writer uses some books as data source having related with topic *researched*." (Volume 7 issue 2)
- "The most widely *researched* of allforms of folkloreby folklore experts is the prose narrative." (Volume 8)
- "The study of Madurese has been *researched* by various disciplines through different topics and discussions." (Volume 9)

in data set 4, the word "*researched*" occurs in the articles taking forms of passive verb and adjective in those sentences. The word *"researched"* in the provided statements indicates the past action of conducting research on a particular topic or subject. It signifies that scholarly investigations have been carried out, involving various methods and approaches to explore and gather information related to the topic under study.

Data set 5

- "These two *researches* are different to this article, particularly in the way this article tries to depict the underlying ideology practiced by the authoritative power in the novel." (Volume 8)
- "They need to pass the seminar proposal or comprehension examination before continuing their *researches*." (Volume 10)
- "The fourth section reviews the current *researches* and any relevant literature on MALL in teaching listening." (Volume 11)

In data set 5, the word "*researches*" occurs in articles taking forms of plural noun either as subject, object or noun phrase in those sentences. The word *"researches"* in the provided statements refers to multiple instances or individual projects of conducting research. It indicates that separate studies or investigations

have been undertaken by different individuals or groups, focusing on specific

topics or areas of inquiry.

2. "Analyse"

In this study, the word "analyse" and its families occur 770 times in the

data. The words associated with "analysis" according to Oxford Dictionary are (1)

analysis as singular noun, (2) analyses as plural noun, (3) analyse as verb British

English version, (4) analyze as verb American English version, (5) analyzed as

American version of past verb, (6) analyzing as present participle of verb form,

(7) analyzing as gerund. In this data, the researcher found the words associated

with "analyse" taking the form of *analyse, analysed, analyses, analysing,*

analysis, analyst, analysts, analytic, analytical, analytically, analyze, analyzed,

analyzes, and analyzing. The following is the explanation data sample:

Data set 6

- "In addition, the writer also uses the concept of discourse ***analyses*** of advertising by Guy Cook (2001)." (Volume 7 issue 2)
- "The careful ***analyses*** in terms of CDA can help social media consumers raise their awareness of how news is manipulated in social media as the media discourse." (Volume 11)
- "His research ***analyses*** Mohja Kahf's The Girl in the Tangerine Scarf (2006) and Diana Abu-Jaber's Crescent (2003)." (Volume 12)
- "His ***analysis*** will hopefully presents information and knowledge of passive aggressiveness that is usually difficult to notice." (Volume 8)
- "He also draws some general conclusions about LL as it applies to the ***analysis*** of visual signs in urban areas." (Volume 9)
- "Discourse ***analysis*** process considers the context of social situations, social practices, and intertextuality of the process of mutual influence (dialectic) between language and society." (Volume 10)

In data set 6, the word "*analyses*" occurs in the articles taking forms of plural

noun and present singular verb either as subject, object or active verb. While the

word "*analysis*" occurs in the articles taking forms of singular noun either as noun

phrase, object or subject. *"Analysis"* refers to the systematic examination and study

of a subject, breaking it down into parts and drawing insights or conclusions.

"Analyses" is the plural form, representing multiple instances or different

approaches of conducting such examinations. They both involve detailed

evaluation and result in the generation of knowledge.

Data set 7

- "This phase is emphasized to ***analyze*** the production of discourse and the social ideologies embedded in every interaction." (Volume 8)
- "His paper attempts to ***analyze*** the death instint manifested through passive aggresivenessshown by Bartleby and its effect toward his surroundings." (Volume 8)
- "Hidayatillah (2017) supposes to ***analyze*** another tradition of Madurese in the form of house design." (Volume 9)
- "At this level, researchers ***analyze*** and explain the relationship between tendencies in the text, complexity in the practice of discourse, and also the processes in social change." (Volume 10)

In the data set 7, the word "*analyze*" occurs in the articles taking forms of

American English verb style either as infinitive verb or present verb. The word

"analyze" in the provided statements refers to the process of examining,

investigating, and scrutinizing a subject or topic in a systematic and detailed

manner. It involves breaking down the subject into its constituent parts, evaluating

its components, and drawing conclusions or insights based on the findings.

Overall, *"analyze"* signifies the process of careful examination and evaluation to

gain a deeper understanding of a subject or topic.

Data set 8

- "In this age of environmental crisis it is enlightening to ***analyse*** the works of this well-known writer against the backdrop of eco-critical theories which take an earthcentred approach to literary studies." (Volume 8)

- "They did not only ***analyse*** the visible signs but also the architecture of the buildings." (Volume 9)
- "In this section, the writer ***analysed*** what it should be looked for at each of the three levels, using data from the beauty product advertisements." (Volume 7 issue 2)

In the data set 8, the word "*analyse*" occurs in the articles taking form of British English verb style either as infinitive verb or present verb. While the word "*analysed*" occurs in the articles taking form of British English verb style as past verb. *"Analyse"* means closely examining and evaluating a subject, while *"analysed"* indicates that the examination has already taken place. It involves breaking down the subject, studying its components, and drawing conclusions or insights. Overall, *"analyse"* and *"analysed"* emphasize the process of careful examination to gain understanding and draw meaningful conclusions.

Data set 9
- "Based on the beauty product advertisement ***analyzed***, the use of direct address is shown below." (Volume 7 issue 2)
- "The criminals' disagreement utterances were ***analyzed*** to find kinds of disagreeing strategies applied by male and female criminals." (Volume 8)
- "In addition to identify the kinds of languages displayed in Sidoarjo LL, the functions of the signs are then ***analyzed***." (Volume 9)
- "According to Fairclough (1989), CDA is a form of research that ***analyzes*** the relationships between discourse, society, power and ideology." (Volume 7 issue 2)
- "One of the studies that ***analyzes*** Taneyan Lanjheng is Hidayatillah (2017)." (Volume 9)

In the data set 9, the word "*analyzed*" occurs in the articles taking form of past verb American style either as past participle verb, or past verb. While the word "*analyzes*" taking form of present verb American style. *"Analyzed"* is the past tense or past participle form of "analyze," indicating a completed examination or evaluation. It suggests a thorough examination of data or discourse to derive

conclusions or insights. *"Analyzes"* is the present tense form, indicating an ongoing

or upcoming examination or evaluation to gain insights and understanding.

Data set 10

- "Based on the story of the novel, the writer is interested in ***analyzing*** one of the
  characters in novel that is Pilar." (Volume 7 issue 2)
- "Therefore, the researcher spent a great deal of her time to read and understand
  the related theories and concepts before collecting and ***analyzing*** the data."
  (Volume 8)
- "Generally, this study also aims at ***analyzing*** the language used in the
  construction of the personal identity of the main character in Nathaniel
  Hawthorne's Dr. Heidegger's Experiment." (Volume 9)
- "They transferred into a new, inexperienced world when they start the research
  journey, by jumping into and ***analyzing*** unanswered question research."
  (Volume 10)

In the data set 21, the word "*analyzing*" occurs in the articles taking form as a

gerund, which functions as a noun. In all three instances, "analyzing" functions as a

noun representing the action or process of examining and interpreting something. It

provided statements refers to the act of closely examining, studying, and evaluating

a specific subject, character, language, or data. It involves breaking down the

subject or data, scrutinizing its components, and drawing conclusions or insights

based on the examination.

Data set 11

- "The descriptive-***analytic*** method is used in analyzing the woman image of beauty
  in Twilight." (Volume 7 issue 2)
- "A descriptive-***analytic*** method of research was utilized in this study, and the data
  were selected from the (full) text of the DWJE." (Volume 10)
- "A descriptive-***analytic*** method of research was utilized in this study, and the
  corpus was analyzed in order to see whether the discourse of the two social
  media editorials promotes any ideological orientations towards the content of the
  news on the historical presence of Portuguese (for more than four centuries) and
  Indonesia (for about 24 years) in Timor Leste (also known as East Timor)."
  (Volume 11)

In data set 11, the word "*analytic*" taking form of adjective to describe the method or approach used in research and analysis. The word "analytic" in the provided statements refers to a methodological approach or perspective that involves careful examination, evaluation, and interpretation of data or information. It suggests a systematic and detailed analysis that aims to uncover patterns, relationships, or insights within the subject or data under study.

Data set 12

- "The general purpose of the three-dimensional model is to provide an ***analytical*** framework for discourse analysis." (Volume 7 issue 2)
- "CDA is a type of ***analytical*** research on discourses which basically looked at how social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk within social and political context." (Volume 10)
- "In his three dimensional model, Fairclough (1989) distinguishes textual, discursive, and social features as three levels that can be ***analytically*** separated." (Volume 7 issue 2)

In data set 12, the word *"analytical"* occurs in the articles taking form of adjective. In this sentence, *"analytical"* is used as an adjective modifying the noun *"framework"* and *"research"*. While the word *"analytically"* taking form of adverb modifying the verb *"separated"*. Overall, *"analytical"* and *"analytically"* emphasize the systematic, methodical, and detailed nature of the examination, evaluation, and interpretation process within the respective contexts of the statements.

Data set 13

- "This issue becomes one aspect of spoken interactions that has been examined by conversation ***analyst***." (Volume 7 issue 2)

- "In this case, the *analyst* explains the results of the interpretation by referring to socio-cultural conditions." (Volume 10)
- "Many CDA *analysts* argue that the literature on media discourse and CDA indicate some studies about media discourse ranging from ethnicity and racism in the media, power relations, and hegemony in the media." (Volume 11)
- "Although many news writers claim that in the process of news writing one should be as objective as possible, neutrality has always remained a question to discourse *analysts*." (Volume 11)

In data set 13, occurring in the articles. In this context the word "analyst" as singular form and "analysts" as plural form word. In the provided statements refers to an individual who engages in the examination, interpretation, and evaluation of specific aspects within a given context. An analyst is someone who systematically studies and scrutinizes data, interactions, discourse, or socio-cultural conditions to derive insights, explanations, or conclusions.

3. **"Text"**

In this study, the word "text" and its families were found to occur 655 times in the data. The words associated with "text" according to Oxford Dictionary are (1) text as singular noun or base verb, (2) texts as plural noun, (3) textual as adjective, (4) subtext as singular noun, (5) subtexts as plural noun. In the data, the researcher found the word "text" taking form as text, texts, textual either as singular, plural noun or as adjective. The following is the explanation of data sample:

Data set 14

- "Grammar includes the structure of words, phrases, clauses, and sentences, right up to the structure of whole *texts*." (Volume 12)
- "It is these elements that play a major role in building the intertext relation of the two *texts*." (Volume 11)
- "This means how we use language, how we communicate in a social situation, and how language works in a *text* or what we do with that language." (Volume 12)

- "In this target *text*, the setting of the story characterizes the condition of American society in 1962." (Volume 11)
- "In the analyzed ten news *texts*, *text* production consists of two main parties; namely the supporters of the inclusion of IF in ID card (E-KTP)." (Volume 10)

In the data set 14, the word "text" occurs in the articles taking the forms of singular noun while the word "texts" taking forms as plural noun. Both of them either as noun clause, object, subject, or noun phrase. Overall, "text" and "texts" represent written or spoken language compositions that can be studied, analyzed, and interpreted within the respective contexts of the statements.

Data set 15

- "The *textual* function is organized by the point of departure- theme and landing point – rheme is the basic *textual* resources in lexicogrammar." (Volume 12)
- "Halliday found that language is organized into three functions: ideational function, interpersonal function, and *textual* function." (Volume 12)
- "The primary data for the analysis were taken from the *textual* and pictorial elements of X-Men." (Volume 11)
- "The quoted text of the news provides *textual* evidence that the use of modality is one vehicle or linguistic proof for building ideology because modalities show subjective or collective attitudes or assessments of the production of news texts on an event." (Volume 10)

In the data set 15, the word "textual" occurs in the articles taking forms of adjectives either as subject, object, adjective of noun phrase. The word "textual" in the provided statements refers to something that relates to or is associated with a text or written material. It denotes aspects, elements, or functions that are specifically connected to the structure, organization, or content of a text.

4. "Identify"

In this study, the word "Identify" and its families were found to occur 616 times in the data. The words associated with "identify" according to Oxford

Dictionary are (1) identity as singular noun, (2) identities as plural noun, (3)

identify as base verb, (4) identifies as present verb), (5) identified as past or past

participle verb. In this data, the researcher found the word associated with

"identify" such as identify, identifies, identified, identity, identities as noun and

verb. The following is the explanation sample data:

Data set 16

- "He expert of Islamic values *identified* some good points in terms of topics
  presented on the draft." (Volume 10)
- "tThe researcher *identified* discourse productivity produced by each camp."
  (Volume 10)
- "The use of relational process indicates that Trump and the American people are
  identified as one nation who have the same pain, dreams, and also success."
  (Volume 9)
- "In his campaign speech, by using relational process Hillary Clinton was
  identified as a bad person such as the greatest liars and disaster." (Volume 9)

In the data set 16, the word *"identified"* in the articles taking forms of past or

past participle verb either as passive past verb or past active verb. The word

"identified" in the provided statements refers to the act of recognizing, determining,

or attributing certain characteristics, qualities, or associations to someone or

something. It implies the process of observing, analyzing, and making a judgment

or assessment based on specific criteria or evidence.

Data set 17

- "the researchers *identify* the use of words and sentences using phonological and
  grammatical knowledge." (Volume 10)
- "The use of Indonesian in the advertisement can *identify* the owner of the shops.
  Scollon and Scollon (2003) distinguish between symbolic and indexical
  functions of bilingual signs." (Volume 9)
- "This study examined how the main character in Nathaniel Hawthorne's Dr.
  Heidegger's Experiment *identifies* his social interaction through his language
  use." (Volume 9)

- "As such, this carries a crucial sociosymbolic importance as it actually ***identifies*** and thus serves as the emblem of societies, communities, and regions (Hult, 2009: 90)." (Volume 9)

In the data set 17, the word *"identify"* in the articles taking forms of base verb as present verb. Meanwhile the word *"identifies"* in the articles taking forms of present singular verb. Overall, *"identifies"* and *"identify"* convey the action of recognizing, determining, or establishing specific qualities, characteristics, or associations within the respective contexts of the statements.

Data set 18

- "Text analysis includes three functions, representation functions, relationships, and ***identities***." (Volume 10)
- "Ideologies usually control the thoughts of a social group which then represent the basic social characteristics of a group based on their ***identities***, goals, norms, values, positions, and resources." (Volume 10)
- "Participant of the context shows who is involved in the event and its ***identity*** (ministers, mass organization figures, people's representatives, etc.)." (Volume 10)
- "Nurti (2017) says that food and cuisine are regarded as forming an ethnic ***identity*** with a typical taste that is different from other ethnicities in other areas." (Volume 9)

In the data set 18, the word *"identities"* found in the articles taking the forms of plural noun as object. Meanwhile the word *"identity"* in the articles taking the forms of singular noun either as object or noun phrase. Overall, *"identities"* and *"identity"* refer to the defining characteristics, qualities, or affiliations that shape individuals or groups, whether in the context of text analysis, social characteristics, participation in events, or ethnic affiliations.

5. **"Strategy"**

In the study, the word "strategy" and its families were found to occur 568 times in the data. The words associated with "strategy" according to Oxford

Dictionary are (1) *strategy* as singular noun, (2) *strategies* as plural noun, (3)

strategic as adjective, (4) *strategically* as adverb, (5) *strategist* as singular noun,

(6) *strategists* as plural noun. In the study, the researcher found words in the

articles taking the forms of *strategy, strategies, strategic and strategically* either

as singular, plural pure nouns, adjective and adverb. The following is the

explanation of data sample:

Data set 19

- "Therefore, this research implies that utilizing story mapping *strategy* can
  improve students' reading comprehension in finding main idea and classroom
  climate." (Volume 9)
- "How the ten hotel homepages employ a particular evaluative *strategy*, the result
  of attitudinal appraisal classification based on the attitudinal appraisal theory is
  described in table." (Volume 10)
- "These *strategies* can be seen as an attempt to negotiate on the minority's part to
  adapt themselves to the society where they live." (Volume 11)
- "Baresove (2008) explored politeness *strategies* in two different cultures,
  American and Japanese, in delivering rejections in letters." (Volume 12)

in the data set 4, we can see "strategy" occurs in articles taking the forms of

singular noun either as object or noun phrase. While "strategies" occurs in articles

taking forms of plural noun either as subject, object or noun phrase. Overall,

"strategy" represents a planned course of action, while "strategies" refers to

multiple planned approaches or methods used to achieve specific goals or

objectives within the respective contexts of the statements.

Data set 20

- "Papua has been a *strategic* target of the USA to conquer because Papua is well
  known for its wealth of natural resources such as gold, coal, and copper."
  (Volume 11)
- "It was then concluded that the *strategic* use of language by professors allowed
  them to negotiate solidarity with students to promote a fairly harmonious and
  democratic relation." (Volume 11)

- "PR's statements were constructed *strategically* and packed in a political discourse designed to alter the American people's state of isolationist minds into a state of patriotic defenders who saw that JE was "the real enemy" who was putting the American land into grave danger." (Volume 10)
- "Harris et al. (1975) said that giving those signals' function is not only as constructing and consolidating social relations but as *strategic* mechanisms for creating transitions into and out of transactional talk." (Volume 7 issue 2)

In the data set 20, we can see *"strategic"* occurs in articles taking forms of adjective as object in the sentence. Meanwhile *"strategically"* taking form of adverb of the verb in the sentence. *"Strategic"* refers to planned, calculated, and goal-oriented actions. *"Strategically"* indicates the intentional and deliberate approach in executing these actions.

6. "Process"

In the study, the word "process" and its families were found to occurs 565 times in the data. The words associated with "process" according to Oxford Dictionary are (1) process as singular noun, (2) process as base verb, (3) processes as plural noun or present verb (4) processed as past or participle verb, (5) processing as noun or gerund. In this study the researcher found some words associated with "process" taking the forms of process, processed, processes and processing either as active or passive verb, adjective, singular or plural noun and gerund. The following is the explanation of the data sample:

Data set 21

- "The data collected were *processed* using AntConc." (Volume 12)
- "Employing these stages, the human language is *processed* in the brain to produce and acquire a proper language." (Volume 11)
- "These behaviors can emerge in the institutional speech in classroom discourse, the area where linguistic devices are *processed* and enacted within the classroom." (Volume 11)

- "Van Dijk (1987, pp. 161-196) also claims that models represent personal experiences such as particular information about each event or action which is *processed* in short-term memory." (Volume 11)

   In the data set 21, the word *"processed"* occurs in the articles taking the forms of passive verb either as past passive verb or present passive verb. The word "processed" in the provided statements refers to the action of handling, treating, or dealing with data, language, behaviors, or information in a specific manner or through a defined procedure. It involves taking input, applying operations or transformations, and generating output or results.

Data set 22

- "It was expected that these *processes* could be attained through the instruments." (Volume 12)
- "Additionally, the flow of events or the *processes* of doing, acting, and events in the advertisements represent Muslim women as having self-reliance, persistence, selfawareness, integrity, and independence as the attributes assigned to the Muslim women in veils." (Volume 12)
- "It implied that reading consists of two related *processes*: word recognition and comprehension." (Volume 10)
- "His chapter analyzes the growth and passage of the researchers during the *process* of their research journey based on their written experiences." (Volume 10)
- "Many prohibitions on the use of the niqab are considered a barrier to the communication *process,* but the pandemic has refuted, requiring masks." (Volume 12)
- "It takes Kimberly a moment to *process* what the barista just said and she shakes her head, she starts to say, but the barista has already moved away to grab the muffin out of the case." (Volume 9)

   In the data set 22, the word *"processes"* occurs in the articles taking the form of plural noun either as object, noun phrase or subject in those sentences. While the word "*process"* in the articles taking the forms of base verb or singular noun either as subject, infinitive verb or object. *"Processes"* refers to a series of interconnected actions or steps, while *"process"* represents a single action or step within a larger

sequence. Overall, *"processes"* and *"process"* signify a series of actions, steps, or operations involved in achieving specific outcomes or understanding within the respective contexts of the statements.

Data set 23

- "Thus, the emergence of hesitation phenomenon in English conversation occurs spontaneously and naturally due to the reason that the ***processing*** of language in the brain is not as easy as it is produced directly through the utterances of words and sentences." (Volume 11)
- "Cowles (2011) points out that some minor delays during language ***processing*** are thought to have profound consequences." (Volume 11)
- "In ***processing*** the moose meat, McCandless had found any difficulty as shown in the part of the story below which happened in sequence 17c." (Volume 10)

In the data set 23, the word *"processing"* occurs in the articles taking the form noun or gerund either as noun clause, noun phrase, or gerund. The word *"processing"* in the provided statements refers to the cognitive or mental activity of handling, understanding, or working with language, information, or specific tasks. It implies the mental operations or procedures involved in dealing with or comprehending something.

7. "Culture"

In this study, the word "culture" and its families were found to occur 524 times in data. The words associated with "culture" according to Oxford Dictionary are (1) culture as pure or singular noun, (2) cultures as plural noun, (3) cultural as adjective, (4) culturally as adverb, (5) cultured as adjective, (6) intercultural as adjective, (7) subculture as singular noun, (8) subcultures as plural noun, and (9) uncultured as adjective. In the data the researcher found several words associated

with "culture" such as culture, cultures, cultural, culturally as adjective, noun and

adverb. The following is explanation of data sample:

Data set 24

- "Prior to that time, *culture* was associated with art, literature, and classical music." (Volume 7 issue 2)
- "Hence, there is still gender inequality and women oppression in Madurese *culture* reflected in the short story." (Volume 8)
- "They are easy to be influenced or shaped by other *cultures*." (Volume 8)
- "African *cultures* were disparaged as barbaric, primitive, irrational, and debased" (Anyaehie, 2013, p. 151)." (Volume 12)

In data set 24, the word *"culture"* occurs in the articles taking form of singular

noun either as subject, object, or noun phrase. While the word *"cultures"* taking

form of plural noun either as subject, object, or noun phrase. *"culture"* refers to the

shared beliefs, customs, values, practices, and behaviors of a group or society,

while *"cultures"* represents multiple groups or societies with their unique cultural

characteristics.

Data set 25

- "This research relates to *cultural* studies since the literary work is popular literature." (Volume 7 issue 2)
- "He attempts to understand whether, in a world driven by the persistent need to develop, update, upgrade and relentlessly move forward, there is still a place for *cultural* memory." (Volume 8)
- "Finally, movie is considered useful to deliver message of *cultural* and social issues in the society." (Volume 9)
- "Later on, when she graduated and started working as a lecturer at a University, she already got used to the "feminine" norms of her society, even state that "For me, this *cultural* norm is a blessing, as I would have been very lonely living away from my family". (Volume 10)

In the data set 25, the word *"cultural"* occurs in the articles taking form of

adjective either as subject or object in the sentences. The word *"cultural"* in the

provided statements refers to aspects related to culture or the shared beliefs,

customs, values, practices, and behaviors of a particular group or society. It pertains

to the social and societal context, traditions, and norms that shape individuals'

identities and experiences.

Data set 26

- "In understanding the tradition of Toron that has been *culturally* widespread,
Djakfar intends to overview the relationships among religious value, culture, and
business ethos within Madurese who is known as hard-working community."
(Volume 9)
- "It has been *culturally* constructed in the society that the marginalized struggle
upon the recognition from the others to claim their worth (Fanon, 2008)."
(Volume 11)
- Politeness is then a *culturally* defined phenomenon that applies good manners and
behavior intended to save the speaker's or the addresses' face (Makejeva, 2017)
(Volume 12)

In the data set 26, the word *"culturally"* occurs in the articles taking form of

adverb of the passive verb in the sentences. The word *"culturally"* in the provided

statements refers to something that is influenced or shaped by the culture of a

particular group or society. It indicates that certain ideas, values, behaviors, or

phenomena are understood or interpreted within a cultural context.

8. **"Data"**

In the study, the word "Data" and its family were found to occur 507 times

in the data, the words associated with "data" according to Oxford Dictionary are

(1) data as plural noun (2) datum as singular noun.in this study, the researcher

found the word "data" only in the form of plural noun and singular noun. The

following sample of data presented below:

Data set 27

- The interaction of Aria and Peregrine with natural environment will be taken as the ***data*** obtained from the novel. (Volume 8)
- In this step, the researcher identified the ***data*** whether they contain error or not. (Volume 9)
- Very little ***data*** indicate it functions to accentuate personal and social identity as was found by Luciana (2006). (Volume 10)
- The ***data*** have revealed that the students who are from Java prefer using negative politeness strategies compared to the ones from outside Java. (Volume 12)

In the data set 27, the word *"data"* occurs in articles taking forms of plural noun either as subject, object, or noun phrase. The word *"data"* in the provided statements refers to factual information or evidence that is collected, observed, or obtained from a specific source or context. It represents the raw or processed information that serves as the basis for analysis, interpretation, or study.

9. "Function"

In this study, the word "function" and its families occur 420 times in the data. The words associated with "function" according to Oxford Dictionary are (1) function as singular noun, (2) functions as plural noun, (3) function as base verb, (4) functional as adjective, (5) functioning as gerund or present participle verb, (6) functionally as adverb. But, in the data the researcher found the word associated with "function" taking forms of function, functions, functioning, functioned, functional, functionally. The following is the data explanation:

Data set 28

- Harris et al. (1975) said that giving those signals' ***function*** is not only as constructing and consolidating social relations but as strategic mechanisms for creating transitions into and out of transactional talk. (Volume 7 issue 2)
- The ***function*** of response items such as 'mm', 'yeah' and 'okay' perform are also influenced by the place and timing of the utterance, or the context. (Volume 8)
- The labels are given by self and others to Dr. Heidegger also have other ***functions*** besides to construct his identity as the main character in the story. (Volume 8)

- There are six ***functions*** of signs available in public places in Sidoarjo. (Volume 9)

In data set 28, the word *"function"* as singular noun occurring in the articles. The word *"function"* in the provided statements refers to the purpose, role, or intended use of something within a specific context. It indicates the particular role or contribution of a certain element or action in achieving a desired outcome or serving a specific purpose. Similarly, the word *"functions"* is the plural form of *"function"* and indicates multiple roles or purposes within a given context. It suggests that there are multiple intended uses or contributions of certain elements or actions in achieving specific outcomes.

Data set 29

- Hall (1998:225; 1996:73) states that representation was understood on the basis of the way words ***functioned*** as signs within language. (Volume 9)
- Wijana's (2006) analysis on the use of in students' writing is not merely ***functioned*** as question words, but also as copula, relative clause marker, interclausal and intersentential conjunctions, and other markers which are difficult to identify because they do not have any equivalent in English. (Volume 10)
- The consultation devices might be ***functioned*** to ask cooperation and totally from the I (speakers)-perspective. (Volume 11)

In data set 29, the word *"functioned"* occurring in the articles is used as a past form of verb or potential action, explaining about the action in a certain context. The word *"functioned"* in the provided statements refers to the action or role performed by something in a specific context. It indicates how an element or entity operates or carries out its intended purpose or role within a given system or framework.

Data set 30

- "In other words, the individuals are systematically trained through task of increasing complexity to prepare them to be a ***functioning*** part of community and to be able to effectively contribute to the common good without questioning the classification of job." (Volume 8)
- Savage and Mooney (1979:1) also state, "Reading comprehension is a process of acquiring or deriving meaning and understanding of printed language; involves cognitive ***functioning*** related to what one reads. (Volume 9)
- Syntax, by definition, is a set of rules that governs the arrangement of words to create a structured, well-***functioning***, and understandable sentence. (Volume 10)

In data set 30, the word "functioning" occurring in the articles is utilized as an adjective to explain something that is processing or working efficiently. The word "functioning" in the presented statements refers to the state or condition of operating, working, or performing as intended within a given system or context. It signifies the effective and successful execution of tasks, roles, or functions to fulfill a specific purpose or contribute to the overall functioning of a system.

Data set 31

- Systemic ***Functional*** Linguistics Theory is the development of grammar study from sentence-based perspective to more discourse-based perspective. (Volume 9)
- A semantic system is organized into three main ***functional*** components, or "metafunctions". (Volume 9)
- Research collaboration is characterized as a system of research activities by several actors linked ***functionally*** and coordinated to achieve a research objective corresponding to those actors' research objectives or interests. (volume 11)
- Though Tamil has been identified 'structurally' as an SOV language, '***functionally***' its grammar allows people to use the same linguistics elements in different ways which are common, meaningful, and acceptable. (Volume 12)

In data set 31, the word *"functional"* occurring in the articles is utilized as an adjective to describe something that relates to the purpose of the working process analysis. While the word *"functionally"* is utilized as adverb. The word *"functional"* in the presented statements refers to the purpose, role, or practical

aspect of a system or component, emphasizing its ability to perform a specific function. The word *"functionally"* describes the manner in which something operates or is performed, highlighting its functional aspects or characteristics within a given context.

10. "Media"

In this study, the word "media" and its families occur 391 times in the data. The words associated with "Media" according to Oxford Dictionary are (1) media as singular or plural noun. In this data, the researcher found the word "media" in the articles taking form media either as singular or plural noun. The following is the explanation data sample:

Data set 33

- Furthermore, Padmadewi and Artini (2017) investigated the implementation of DI in English class through visual *media* as ASD students tended to be visual learners. (Volume 12)
- The direction of *media* representation of Muslim women is gradually leading to a better direction. (Volume 12)
- The discipline's growth in breadth and depth has also induced studies and discussions on film's relations with other *media* such as television and literary work, particularly novels. (Volume 11)
- In developing English teaching materials, *media* play an important role to support the implementation of teaching materials development. (Volume 10)

In the data set 33, the word "media" occurs in the articles taking forms of singular or plural form either as noun phrase, subject or object. The word "media" in the presented statements refers to various forms of communication channels or platforms that are used for conveying information, messages, or content. It encompasses different mediums such as visual media, television, literary works, and teaching materials. The term "media" highlights the importance of these

channels in disseminating information, facilitating communication, and supporting various activities or purposes within the respective contexts.

4.1.2 Coverage of Academic Vocabulary in Every Volume

In this section the researcher presents the finding of the analysis responding to the second research problem of the study about the coverage of academic words in each volume. The following data explained below:

Table 4.3.   AWL Coverage in Volume 7 – Volume 12

No	Volume	AWL	GSL
1	Vol 7 issue 2	1852 tokens (7,2 %)	21305 tokens (82.50%)
2	Vol 8	3410 tokens (7,75 %)	34724 tokens (78,97%)
3	Vol 9	4379 tokens (8,91 %)	38855 tokens (79,04%)
4	Vol 10	5555 tokens (9,27 %)	47182 tokens (78,71%)
5	Vol 11	7709 tokens (9,55 %)	63393 tokens (78,5%)
6	Vol 12	8492 tokens (10,23 %)	63606 tokens (76,64%)

The table above presents the results of the AWL coverage and GSL coverage analysis across all volumes. It also provides the token counts for each volume. The analysis of AWL coverage reveals a significant pattern observed throughout the volumes. The token counts of AWL consistently demonstrate development, indicating an increased utilization of academic words. This development suggests that the AWL and GSL utilized in the volumes may have evolved over time. Furthermore, the AWL percentages in the corpus, which measure the proportion of AWL tokens to the total, show a rising trend.

The volumes, starting from Volume 7 Issue 2, account for 7.2% of the corpus, with 1,852 tokens. Volume 8 accounts for 7.75% of the corpus, with 3,410 tokens. Volume 9 accounts for 8.91% of the corpus, with 4,379 tokens. Volume

10 accounts for 9.27% of the corpus, with 5,555 tokens. Volume 11 accounts for 9.55% of the corpus, with 7,709 tokens. Lastly, Volume 12 accounts for 10.23% of the corpus, with 8,492 tokens. These findings indicate that although AWL tokens comprise a relatively small portion of the total token count in the corpus, their significance and contribution to the articles are consistently improving. The increasing token counts and percentages emphasize the importance of AWL coverage in each volume.

On the other hand, GSL tokens dominate the token counts across all volumes. With percentages ranging from 76.64% to 82.50%, the GSL category consistently represents the majority of the total tokens. This implies that the majority of tokens in the volumes fall within the scope of the widely used General Service List (GSL). The consistent percentages of GSL throughout the volumes indicate the inclusion of words with a broader, more commonly understood language. This suggests that the volumes contain a substantial amount of general English words accessible to non-native speakers or individuals focusing on common and widely known language usage.

The comparative analysis of the occurrences of Academic Word List (AWL) and General Service List (GSL) tokens reveals that GSL tokens have a higher distribution in the volumes compared to AWL tokens. This dominance of GSL tokens highlights the prevalence of widely understood language patterns and usage. Furthermore, the increasing frequency of AWL tokens indicates a potential expansion in the variety and length of academic words used in the volumes. This signifies progress in the NOBEL Journal Articles, suggesting that the publishers'

editors have made efforts to maximize and maintain the selection of journal

articles' content each year.

Table 4.4 The Top 10 Word Frequency in Each Volume

No	Volume 7	Volume 8	Volume 9	Volume 10	Volume 11	Volume 12
1	**strategy**	feature	process	**research**	**research**	**identify**
2	respond	**research**	error	**text**	**text**	**research**
3	**analyse**	**analyse**	**research**	**analyse**	**analyse**	**strategy**
4	**data**	community	**text**	participate	culture	media
5	**text**	**data**	**analyse**	process	media	**analyse**
6	**research**	**strategy**	**identify**	function	context	**data**
7	consume	**identify**	label	positive	process	function
8	function	contradict	culture	sequence	communicate	concept
9	feature	culture	symbol	technology	**data**	culture
10	**identify**	ideology	concept	lecture	expose	positive

The table above describes selected Volume 7 – Volume 12. Each volume

has a list of words associated with academic words. It is indicated from the table

that from the data above through analysis based on the AWL list. the researcher

presents the top 10 vocabulary that appears based on the highest number of

frequencies. The words presented above are headwords which also contain their

families. the results show some of the academic vocabularies that appear

frequently across volumes. There are several bold words mostly occur in all

volumes such as *analyse*, research, data, strategy and identify. While for the least

words occurrence such as sequence, expose, lecture, positive, context and etc.

In Volume 7, which comprising the vocabulary keywords and the sub-list

such as strategy (2), respond (1), analyse (1), data (1), text (2), research (1),

consume (2), function (1), feature (2), and identify (1), these words are the

vocabulary key of the most occurrence words. the statistic result shows the word types of the GSL coverage is 57,14% of the corpus accounted for 1729 types, while for the AWL is 16,49% of the corpus accounted for 499 types. The headwords for the GSL coverage is 46,21% of the corpus accounted for 889 headwords, while the headwords for the AWL is 14,23% of the corpus accounted for 287 headwords.

In Volume 8, which comprising the vocabulary keywords and the sub-list dispersed in the corpus such as feature (2), research (1), analyse (1), community (2), data (1), strategy (2), identify (1), contradict (8), culture (2), ideology (7). these words are the vocabulary key of the most occurrence words. the statistic result shows the word types of the GSL coverage is 48,2% of the corpus accounted for 1848 word, while for the AWL is 15,71% of the corpus accounted for 809 types. The headword for the GSL is 38,08% of the corpus accounted for 1216 headwords, while the headwords for the AWL is 11,31% accounted for 392 headwords.

In volume 9, which comprising vocabulary keywords and the sub-list dispersed in the corpus such as Process (1), error (4), research (1), text (2), analyse (1), identify (1), label (4), culture (2), symbol (5), and concept (1). The statistic result shows that the word types of the GSL is 47,3% of the corpus accounted for 2583 types, while for the AWL is 15,35 of the corpus accounted for 838 types. The headwords coverage for the GSL is 34,4% of the corpus accounted for 1283 headwords, while for the AWL is 10,94 of the corpus accounted for 408 headwords.

In volume 10, which comprising vocabulary keywords and sub-list dispersed in the corpus such as research (1), text (2), analyse (1), participate (2), process (1), function (1), positive (2), sequence (3), technology (3), lecture (6). The statistic reports that the word types for the GSL coverage is 44,24% of the corpus accounted for 2693 types, while for the AWL coverage is 17,2% of the corpus accounted for 1047 types. The headwords coverage for the GSL is 30,69 of the corpus accounted 1237 headwords, while for the AWL coverage is 11,07% of the corpus accounted for 446 headwords.

In volume 11, which comprising vocabulary keywords and sub-list dispersed in the corpus such as research (1), text (2), analyse (1), culture (2), media (7), context (1), process (1), communicate (4), data (1), expose (5). The statistic reports that the word types for the GSL coverage is 43,11% of the corpus accounted for 3074 types, while for the AWL coverage is 16,98% of the corpus accounted for 1211 types. The headwords coverage for the GSL is 29,01% accounted for 1357 headwords, while for the AWL coverage is 10,16% of the corpus accounted for 475 headwords.

In volume 12, which comprising vocabulary keywords and sub-list dispersed in the corpus such as identify (1), research (1), strategy (2), media (7), analyse (1), data (1), function (1), concept (1), culture (2), positive (2). The statistic reports that the word types for the GSL coverage is 41,58% of the corpus accounted for 2877 types, while for the AWL coverage is 16,74% of the corpus accounted for 1158 types. The headwords coverage for the GSL is 27,84% of the corpus accounted for 1295 headwords, while for the AWL coverage is 10,15% of the corpus accounted for 472 headwords.

4.2   Discussion

In this section, the researcher will analyze and discuss the findings of the research, comparing them to previous studies and relevant theoretical frameworks. The primary focus is on addressing the first research question and examining the lexical composition of the NOBEL journal articles in relation to the General Service List (GSL) and Academic Word List (AWL) base lists. Upon analyzing the results, it was observed that the 1,000 most frequently used English words from the GSL accounted for 253,303 tokens, representing approximately 73.95% of the entire corpus. The subsequent 1,000 common words from the GSL constituted 15,662 tokens, equivalent to 4.57% of the corpus. Overall, the coverage provided by the GSL base lists was calculated to be 78.52%.

In addition, the AWL contributed 31,398 tokens or 9.17% of the entire corpus. By merging the GSL and AWL's investigation, the total percentage reached 87.69%. This indicates that 12.31% of the corpus, or 42,179 tokens, were of words outside the scope of the GSL and AWL. The findings will be analyzed and discussed in the context of the relevant literature and theories to provide a full comprehension of the lexical composition of NOBEL journal articles and the extent of coverage provided by the GSL and AWL.

The findings of the currently underway research are similarly to those of previous research on applied linguistics and academic word dispersion. In the current investigation, 9.17% of the analyzed corpus is covered by the AWL. This study confirms previous findings that this academic word list constitutes approximately 10% of the majority of academic words (Coxhead & Byrd, 2007;

Hyland & Tse., 2013; Khani & Tazik, 2013; Chanasattru & Tangkiengsirisin,

2017; Matinparsa et al, 2022).

Table 4.5 AWL Coverage in The Current Research and Previous Researches

Word List	(Hyland & Tse, 2007)	(Khani & Tazik 2013)	Chanasattru & Tangkiengsirisin (2017)	Matinparsa et al. (2022)	This study
GSL	74	76.40	-	73.78	78.52
AWL	10.6	11.96	13.86	11.46	9.17
GSL + AWL	84.7	88	-	85.24	87,75

Comparable investigations have previously examined the dispersion of

academic terms in articles published in academic journals. Hyland and Tse (2007)

examined the dispersion of the academic term in a corpus containing numerous

academic disciplines, including Engineering, Social Science, and Science.

According to Hyland and Tse (2007), the AWL represented 10.60% of the total

corpus. Indicating its presence in a substantial portion of the academic texts

analyzed. In addition, the cumulative representation given by the General Service

List (GSL) approached 85 percent of the corpus, indicating that the GSL included

a substantial portion of the vocabulary encountered in academic articles across the

three disciplines. These findings from Hyland and Tse's (2007) study contribute to

the existing literature on the distribution of academic words, shedding light on the

prevalence of the AWL and the coverage provided by the GSL within the specific

disciplinary context of Engineering, Social Science, and Science.

Khani and Tazik (2013) confirmed in their research that the total number

of tokens in their corpus was reported as 1,553,450 running words. The corpus

consisted of 32,479 word types and 2,409 word families. The General Service List

(GSL) provided a coverage of 76.4% of the total corpus, amounting to 1,165,088 tokens. The Academic Word List (AWL) coverage accounted for 11.96% of the tokens, corresponding to 194,355 running words. Similarly, Chanasattru and Tangkiengsirisin (2017) examined the word list distribution in social science research articles. They utilized the New General Service List (NGSL) for the distribution of common English word lists. The NGSL coverage accounted for 303,425 running words, representing 73.2% of the entire corpus. The AWL coverage, on the other hand, accounted for 57,445 running words, constituting 13.86% of the corpus.

Others study, Matinparsa et al. (2022) proved in their research examined the distribution of articles using the GSL and AWL word lists. The GSL coverage accounted for 35,626,816 running words, representing 73.78% of the corpus. The AWL coverage, in contrast, accounted for 553,632 running words, comprising 11.46% of the corpus. These findings from Khani and Tazik (2013), Chanasattru and Tangkiengsirisin (2017), and Matinparsa et al. (2022) contribute to our understanding of the distribution of academic words in research articles. They highlight the coverage provided by the GSL and AWL in different corpora, indicating the prevalence and significance of these word lists in the analyzed academic contexts.

It can be inferred from the previous studies that the overall comparing researches indicate that most previous research achieved above 10% of the AWL. About some data are journal articles from international reputable publisher. For example the study conducted by (Chanasattru & Tangkiengsirisin, 2017) that they employed the data from international reputable journal articles namely

(ScienceDirect.com | Science, Health and Medical Journals, Full Text Articles and Books., n.d.) is the foremost global platform providing access to extensive scientific such as journals, books, and articles with editors team who maintain a collective competence of over 140 years in reviewing and verifying scientific knowledge. Numerous researchers published their research in ScienceDirect are international researchers who have the competence in writing academic papers. Unlike in NOBEL which the researchers mostly from local people.

In their study on academic vocabulary in applied linguistics, Matinparsa et al. (2022) utilized the data obtained from journal articles in the field. The researchers relied on reputable journal ranking systems, such as Scimago Journal ranking, which assesses the quality and impact of journals in the domain of linguistics and language. Their methodology involved referencing well-established scientific databases like Web of Science and Scopus. Furthermore, the researchers adhered to the recommendations provided by ten esteemed university professors specializing in applied linguistics.

Both previous researches given the resulting data that indicates better academic word coverage in their analyzed articles. Unlike the current study, the overall coverage of the entire corpus of journal articles published between 2016 (Volume 7 issue 2) – 2021 (Volume 12) shows below the asverage 10% of the AWL standard list. The Nobel Journal articles have not maximized the usage of academics in the articles. But the researcher found that responding to the second research question finding. The result showed interesting pattern of increasing. there is significant increases in academic word coverage from year to year from volume 7 issue 2 to volume 12 which published between 2016 - 2021. This

indicates that the editor and publisher have striven to improve the quality of the journal articles to be more comprehensive in the following volumes. This suggests that future journals can achieve significant improvements if their editorial teams validate articles by carefully considering the appropriate usage of academic words within the articles.
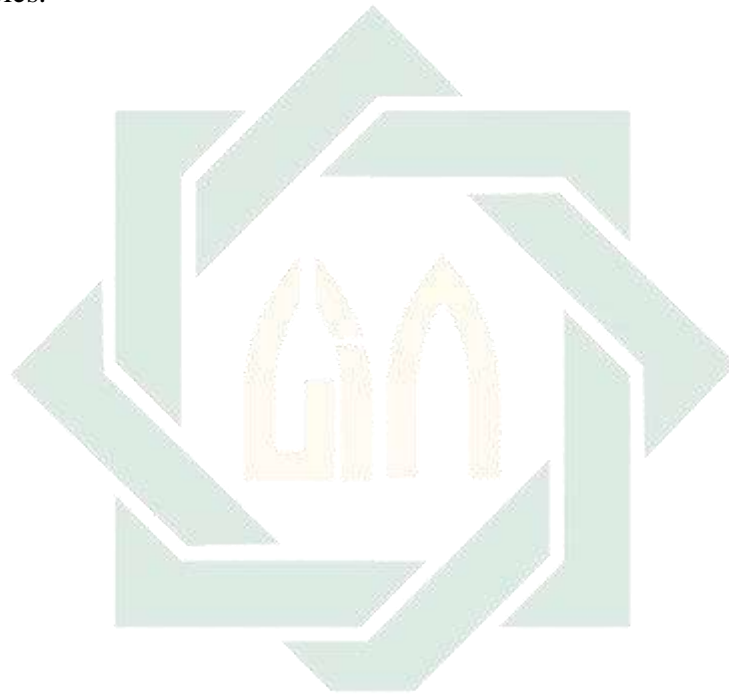
CHAPTER V

CONCLUSIONS AND SUGGESTIONS

This chapter provides the conclusion of the current research about the examined data of the distribution of academic words in NOBEL Journal Articles and also suggestions for future research relevant to the distribution of academic words in journal articles.

5.1 Conclusions

The current study examines the distribution of academic words in NOBEL Journal Articles. The data analyses were conducted into two corresponding to the first research problem: the distribution of academic words in entire journal articles corpus. The second research problem: the distribution of academic word analyzed in every volume of journals. The analysis was conducted by utilizing the software analysis called AntWordProfiler to calculate the data and AWL and GSL list as the reference list. The outcome of statistic shows that the coverage of academic words in entire articles reach 9,17% of the corpus and added with 78.52% of the GSL coverage. The top ten most occurrence words are listed such as *Research, analyse, text, identify, strategy, process, culture, data, function*, and *media.*

For the second problem of the study. The researcher discovered the statistic data every volume from volume 7 issue 2 – volume 12. The resulting analysis for the AWL coverage in volume 7 is 7,2% of the corpus while for the GSL coverage is 82,50%. The AWL coverage in volume 8 is 7,75% of the corpus while the GSL coverage is 78,97. The AWL coverage in volume 9 is 8,91% of the corpus while the GSL coverage is 79,04%. The AWL coverage in volume 10 is 9,27% of the

59

corpus while the GSL coverage is 78,71%. The AWL coverage in volume 11 is 9,55% of the corpus while the GSL coverage is 78,5%. The AWL coverage in volume 12 is 10,23% of the corpus while the GSL is 76,64%. From the statistic shows significant development of higher coverage of academic words from year to year. This indicates better improvement through the following years for the future journals if the editorial teams validate the articles by considering the selection of academic word usage in the articles.

5.2 Suggestions

However, it is important to acknowledge that this research study does have certain limitations. One limitation pertains to the selection of the reference word list, specifically Coxhead's Academic Word List (AWL), which is derived from a corpus that only encompasses 570 word families. Consequently, the analysis of the data is restricted due to the limited coverage provided by this word list. It is worth noting that there are other reference word lists available. Browne, et al (2022) developed the NGSL (New General Service List) 1.2 is a collection of 2809 essential words in general English that are crucial for second language learners in their daily lives. While this word list represents less than 10% of the vocabulary mastered by native speaker college graduates, it provides an average coverage of 92% for most general English texts and even higher coverage in various contexts. For instance, it covers 93% of the vocabulary used in the Harry Potter series, 94% of the vocabulary found in TOEIC exams, and 95% of the vocabulary used in popular TV shows like Friends.

The NAWL (New Academic Word list) 1.2 is an updated version of Averil Coxhead's Academic Word List (Coxhead, 2000), consisting of 957 essential words for general academic English. It was developed to complement the NGSL 1.2 and serve as a substitute for Coxhead's original list. The need for replacing the original AWL arose due to three reasons. Firstly, the NGSL contains a different number of words compared to the original GSL by West (1953). Secondly, the NGSL Project utilizes modified lexeme sets (or flemmas) as word units, contrasting with the word family approach employed by the AWL. Lastly, the goal was to achieve broader coverage than the list it replaces.

Moreover, the study's inclusion of journal articles was limited to volume 12 published in 2021, while the latest volume, Volume 13 in 2022, has been released by the NOBEL journal. It should be noted that Volume 14 of the NOBEL journal is currently in the process of being published, and its findings may contribute valuable insights that were not considered in this study. Suggestion for future researchers to utilize the updated corpus and reference word list in order to provide the wider scope of corpus for the better word list distribution analysis research.

REFERENCES

Anthony, L. (2022). *AntWordProfiler (Version 2.0.1) [Computer Software].* Tokyo,Japan:WasedaUniversity. Retrieved on February 15, 2023 from https://www.laurenceanthony.net/software

Chanasattru, S., & Tangkiengsirisin, S. (2017). The Word List Distribution in Social Science Research Articles. Arab World English Journal, 8 (4). https://dx.doi.org/10.24093/awej/vol8no4.28

Carter, R., & McCarthy, M. (1988). Vocabulary and Language Teaching. Longman.

Charlie Browne. (n.d.). Free Resources. Retrieved from https://www.charlie-browne.com/free-resources

Chen, Q., & Ge, G. C. (2007, January). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, *26*(4), 502–514. https://doi.org/10.1016/j.esp.2007.04.003

Chomsky, Noam (2000): The Architecture of Language. Oxford: Oxford UniversityPress.

Damico J. S., & Ball, M. J. (2019). *The SAGE Encyclopedia of Human Communication Sciences and Disorders*. https://doi.org/10.4135/9781483380810

Evison, J. (2010). Corpus linguistics: An introduction. Edinburgh University Press.

Gardner, D & Davies, M. (2013). A New Academic Vocabulary List. Oxford University Press.

Hyland, K. and Tse, P. (2007), Is There an "Academic Vocabulary"?. TESOL Quarterly, 41: 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Hoda Matinparsa, Ismail Xodabande, Mohammad Ghafouri et al. Academic Vocabulary in Applied Linguistics Research Articles: A Corpus-based Replication Study, 27 September 2022, PREPRINT (Version 1) available at Research Square https://doi.org/10.21203/rs.3.rs-2092705/v1

Hartono, D. A., & Badara Prima, S. A. (2021). The Correlation Between Indonesian University Students' Receptive Vocabulary Knowledge and Their Reading Comprehension Level | Hartono | Indonesian Journal of Applied Linguistics. https://doi.org/10.17509/ijal.v11i1.34590

Khani, R., & Tazik, K. (2013). Towards the Development of an Academic Word List for Applied Linguistics Research Articles. RELC Journal, 44(2), 209–232. https://doi.org/10.1177/0033688213488432

Mabruroh, N. L. (2020, January 1). I Hate Reading; Revealing Academic Reading Difficulties of Indonesian Learners of English. Retrieved from https://repository.unej.ac.id/handle/123456789/98159

Nation, I. S. P. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N., & McCarthy, M. (Eds.), Vocabulary: Description, Acquisition and Pedagogy (pp. 6-19). Cambridge University Press.

Nation, P. (1990). Teaching and Learning Vocabulary. Heinle & Heinle Publishers.

NOBEL: Journal of Literature and Language Teaching. (2023, April 28). https://jurnalfahum.uinsby.ac.id/index.php/nobel
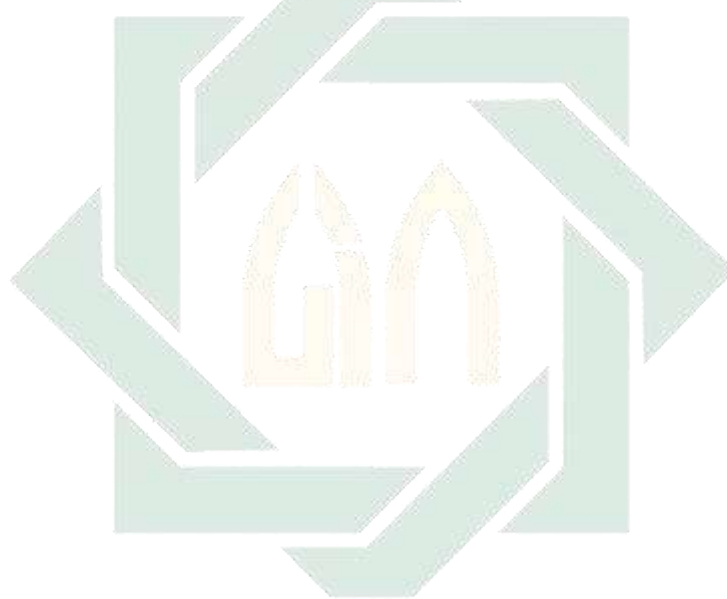
Prasetya, E. P., Ekawati, A. D., Nugraha, D. S., Marzuq, A., & Darlis, T. S. (2020). Corpus Linguistics, Language Corpora and Language Teaching. *English Journal, 14(2)* retrieved on 13 June 2023 from https://doi.org/10.32832/english.v14i2.3845

ScienceDirect.com | Science, health and medical journals, full text articles and books. (n.d.). ScienceDirect.com | Science, Health and Medical Journals, Full Text Articles and Books. https://www.sciencedirect.com

Shaw, P. (1991). Science research students' composing processes. English for Specific Purposes

Thorndike, E. L., & Lorge, I. (1944). The teacher's word book of 30,000 words.

West, M. (1953). A General Service List of English Word