Creating Indonesian Academic Vocabulary by Using the Ant Word Profiler Program to Academic Writing for BIPA Learning

(Penyusunan Daftar Kosakata Akademik Bahasa Indonesia dengan Menggunakan Program *Ant Word Profiler* untuk Penulisan Akademik bagi Pemelajar BIPA)

Dadang Rhubido
Universitas Negeri Surabaya, Indonesia
dadangrhubido@unesa.ac.id

Syamsul Shodiq
Universitas Negeri Surabaya, Indonesia
syamsulsodiq@unesa.ac.id

Prima Vidya Asteria*
Universitas Negeri Surabaya, Indonesia
primaasteria@unesa.ac.id

*Corresponding author: Prima Vidya Asteria | email: primaasteria@unesa.ac.id

Abstract: The list of Indonesian academic vocabulary has never been made. This study aims to create a list of Indonesian academic vocabulary. This study also explains how to create a list of academic vocabulary using the Ant Word Profiler program. The method used in this study is a combined research method (mixed methods), namely the combination of quantitative and qualitative methods. The research data used in this study were taken from 200 scientific articles from accredited national journals covering the fields of social, health, engineering, and life sciences. The research data amounted to 672,787 words. The stages of research data analysis are (1) knowing the list of Indonesian academic vocabulary, (2) making a list of Indonesian academic vocabulary (3) describing the research results. This researh produced a product in the form of a list of Indonesian academic vocabulary consisting of 235 word families. The list of academic vocabulary that has been made in this study can be used by BIPA teachers and students to be more effective and efficient in learning the Indonesian language so that it easier for BIPA students to produce academic texts in Indonesian.

Abstrak: Daftar kosakata akademik Indonesia belum pernah dibuat. Penelitian ini bertujuan untuk membuat daftar kosakata akademik Indonesia. Penelitian ini juga menjelaskan cara membuat daftar kosakata akademik menggunakan program *Ant Word Profiler*. Metode yang digunakan dalam penelitian ini adalah metode penelitian gabungan (*mixed methods*), yaitu kombinasi metode kuantitatif dan kualitatif. Data penelitian yang digunakan dalam penelitian ini diambil dari 200 artikel ilmiah dari jurnal nasional terakreditasi yang meliputi bidang sosial, kesehatan, teknik, dan ilmu hayati. Data penelitian berjumlah 672.787 kata. Tahapan analisis data penelitian adalah (1) mengetahui daftar kosakata akademik bahasa Indonesia, (2) membuat daftar kosakata akademik bahasa Indonesia (3) mendeskripsikan hasil penelitian. Penelitian ini menghasilkan produk berupa daftar kosakata akademik Indonesia yang terdiri dari 250 rumpun kata. Daftar kosakata akademik yang telah dibuat dalam penelitian ini dapat digunakan oleh guru dan siswa BIPA agar lebih efektif dan efisien dalam pembelajaran bahasa Indonesia sehingga memudahkan siswa BIPA dalam menghasilkan teks akademik dalam bahasa Indonesia.

INTRODUCTION

Data shows that BIPA students in the world number 8,950 people in 80 institutions spread across 30 countries (kemendikbud.go.id). They learn Indonesian for various purposes, for example economics, education, politics, etc. The high interest of foreigners to learn Indonesian must be anticipated by preparing teaching methods that are easy to understand (Moeliono et al., 2017). They

consider learning Indonesian as a practical goal. Due to the time limitation, they hope to master Indonesian in a short time indeed. The right approach and learning method are certainly needed. An effective learning method is a method that provides what students need. Read (2020) states that the more vocabulary a person masters, the greater the possibility of that person's success in academics, business and careers (Browne, 2014). In the corpus, the Indonesian vocabulary is lots, coupled with derivative and absorption vocabulary. These vocabularies are not all needed by foreign students in their fields. They only need to learn, remember, and use linguistic forms in accordance with their practical fields.

Vocabulary is all words contained in a language. According to Wilkins, vocabulary is the foundation in language activities (Thornburry, 2022). According to him further, if someone does not master the grammar, then he can still express an intention to the speaker. However, if someone does not master the language vocabulary, then he can not express what he wants to convey to the other person (Cindana, 2022). It illustrates that the mastery of vocabulary is the basis of four language skills, including listening, speaking, reading, and writing.

In second language learning, students should be introduced to a number of vocabularies and their meanings so that gradually they become to understand and master the form of vocabulary and its relationship with the semantic aspects (Yang, 2015; Yusuf 2021). Someone must master the vocabulary and its use both in spoken and written (Darwis & Kamsinah, 2019; Teddlie & Yu, 2007). In another sense, vocabulary can be said as the basic unit of language that must be mastered by someone who wants to communicate with others using verbal language. A person's language ability is influenced by his vocabulary mastery (Putri et al., 2022; Sari et al., 2021; Winarti, 2023). The more vocabulary a person has, the better his language skills will be (Susanto, 2017; Supriadin, 2016).

The use of appropriate learning methods is very important in language learning. Steps and sequences of learning that are coherent and sequential become things that must be considered. The practical goal in language learning must be its main goal, namely to make students are able to communicate with the language they learned (Henilia, 2023). Before learning language widely in the level of sentences and even discourse, someone must learn the smallest units first, namely vocabulary. At present, many experts think that in teaching, educators need to know the minimum number of words that students must master (Nation, 2001). It aims to make educators more precise in providing a number of vocabularies that must be learned by students so that the learning process becomes more effective. For example, for the purpose of scientific writing, students are sufficiently introduced to a number of academic vocabularies which are only a fraction of the total vocabulary of a language (Herpondo et al., 2022; Indhiarti & Chaerunnisa, 2021; Mujianto & Sudjalil, 2021).

Vocabulary has many types according to the realm of its use. Nation classifies vocabulary into (1) high-frequency vocabulary, (2) academic vocabulary, (3) technical vocabulary, and (4) low-frequency vocabulary. Vocabulary that is functionally more used in academic activities is called academic vocabulary (Nation, 2001). Academic vocabulary is composed of a number of technical vocabularies from various disciplines. The academic vocabulary must appear in the range of sub-corporations that represent various disciplines (Lei & Liu, 2016). Similar studies have been made by Coxhead (2000) and Gardner & Davies (2014). An explanation of the research of Coxhead (2000) and Gardner & Davies (2014) is discussed in the literature review section. In addition, a similar study was conducted by Kwary (2013) with the title "Creating and Testing the Indonesian High Frequency Word List". In this study, an Indonesian High Frequency Word List was produced, which consisted of 500 word families. This research is different from this research because this research will create an academic word list (Sutami, 2014).

In addition, Astuti & Aziez (2021) in their research entitled "*Penyusunan Daftar Kosakata Baca Bahasa Indonesia untuk Siswa Sekolah Menengah Pertama (SMP)*" made a Indonesian reading vocabulary list for junior high school students. The research data was taken from the junior high school's textbook of the Kemendikbud in 2016. The objects and results of this research are different from this research. The methods and tools used are also different. Therefore, research on making Indonesian academic vocabulary for foreign speakers needs to be done because this research has never been done.

Based on the descriptions above, this study aims to explain the process of making a Indonesian academic vocabulary list using the Ant Word Profiler program. In addition, the research will also produce a Indonesian academic vocabulary list. Vocabulary has many definitions. However, all definitions refer to the same thing. Webster defines vocabulary as (1) a list of words or phrases that are usually arranged in an orderly alphabetical way as in a dictionary or glossary. (2) All words from the language, and (3) All words used by a particular group of people are recognized and understood by that group of people, even though they do not always use them (Webster's, 2002).

Vocabulary is all the words contained in a language in addition. Vocabulary must be mastered by foreign language learners to be able to acquire language proficiency. Vocabulary is the first step in mastering language skills (Selviana et al., 2020). Vocabulary as a list of words used by someone in books and other language media. In other words, vocabulary is a list of words that contain meaning according to the agreement of a speech community group, especially those contained in textbooks.

From these definitions, it can be concluded that the vocabulary is a list of words contained in a language that is agreed upon and used by a group of speech communities in communication. Of the vocabulary importance, it is no exaggeration if the vocabulary is referred to as the language foundation.

Nation (2001) classifies vocabulary into (1) high-frequency vocabulary, which is the vocabulary that we encounter most often in the text. (2) academic vocabulary, which is the vocabulary used and found in academic texts. This vocabulary is typically found in textbooks. Nation argues that academic words usually cover almost 9% of the words contained in books. (3) technical vocabulary, namely vocabulary that is specifically used in certain domains. This vocabulary group can identify the realm that uses it. According to Nation, technical words cover about 5% of the words contained in the text. (4) Low-frequency vocabulary, which are words that are rarely used in the text so that the scope of word occurrence is very small.

Research on the list of academic vocabulary was done by Averil Coxhead. Coxhead's research produced 570 family words which commonly found in English academic texts. Heretofore, a list of academic words made by Coxhead is considered one of the most well-known lists of general academic words (Valipouri & Nassaji, 2013). Coxhead academic vocabulary list covers up to 10% of the vocabulary covered in academic texts. It means that one out of ten words in an academic text is the academic vocabulary of the text.

The Coxhead academic vocabulary research is intended as a reference for advanced students who want to study in English. The reason for making the academic vocabulary list was because Coxhead realized the students' difficulties in mastering the vocabulary needed to write assignments. Knowledge of academic vocabulary is very important for students, especially for reading and writing. Coxhead academic vocabulary list is intended to make students more effective in learning English especially in the field of scientific writing. Coxhead illustrates that English native speakers at the age of 18 have at least 18,000-20,000 English vocabularies. It certainly requires a long time for English learners. Therefore, the academic vocabulary list was made. Only by mastering 570 academic vocabularies, then English learners are able to write scientific in English.

Coxhead academic vocabulary research utilizes corpus linguistics. Coxhead linguistic data are made with the criteria (1) linguistic data taken from four different faculties, namely the faculty of art, faculty of economics, faculty of law, and faculty of science and technology. (2) The vocabulary appears more than one hundred times in the corpus as a whole. (3) The vocabulary appears at least ten times in each faculty. (4) It is a vocabulary beyond 2000 vocabularies that often appears in the Michael West's General Service List.

In the distribution of linguistic data, Coxhead collected 158 articles from academic journals consisting of 51 academic journal articles from the website, 43 data taken from university textbooks, 83 data from Learned and Scientific from Wellington Corpus of Written English and from The Brown Corpus (Coxhead, 2000). In addition, research data were also taken from 33 chapters from university textbooks, 31 manuscripts from the Learned and Scientific section of Lancaster-Oslo / Bergen (LOB)

Corpus, 13 academic psychology textbooks part of the Micro Concord academic corpus, and from corpus data that taken from 2 universities. The following is a corpus data table used by Coxhead.

Table 1.
Corpus Data Composition By Coxhead (2000)

	Discipline				Total
	Arts	Commerce	Law	Science	
Running words	883,214	879,547	874,723	875,846	351,333
Texts	122	107	72	113	414
Subject areas	Education History Linguistics Philosophy Politics Psychology Sociology	Accounting Economics Finance Industrial relations Management Marketing Public policy	Constitutional Criminal Family and medicolegal International Pure commercial Quasi-commercial Rights and remedies	Biology Chemistry Computer science Geography Geology Mathematics Physics	

Coxhead academic vocabulary list is organized based on word family which consists of parent word and family members. In the study, it exemplified by the word maximize which has family members include inflections of verbs such as maximized and maximizing and noun such as maximum.

Coxhead vocabulary research for several decades became the basic theory in similar studies in several areas. However, in its development, there are also many linguists who criticize the workings of Coxhead which is considered a lot of gaps that must be refined. However, many other linguists also continue to use the Coxhead method. They assume that the Coxhead method is not wrong, but it only needs to be developed and refined following the development of corpus linguistics. In the development of corpus linguistics, many linguists offer a larger corpus database that offers work efficiency. This is influenced by the rapid development of information technology. Research and creating a list of new academic vocabularies by perfecting Coxhead theory was conducted by (Gardner & Davies, 2014). Gardner & Davies (2014) named their research results with the New Academic Vocabulary List which is commonly abbreviated as NAVL.

In that study, Gardner & Davies (2014) processed 120 million words on the new corpus list. The data are greater than what Coxhead did. Moreover, in contrast to what Coxhead mentioned that the list of academic vocabulary is 10% of the vocabulary contained in the text, Garner and Davies state that the list of academic vocabulary they made is approximately 14% of the vocabulary contained in the text. That means that 4% greater than what Coxhead mentioned. Gardner & Davies (2014) questioned Coxhead's work steps regarding the use of family words and the relationship of academic vocabulary to word lists in public services. Gardner & Davies (2014) disagree if the academic vocabulary list is classified based on the words family since it will cause meaning problems. That is because the words family does not consider grammatical aspects of words, such as nouns, verbs, adjectives, adverbs, etc. In addition to the above problem, Gardner and Davies also questioned Coxhead's theory of the exclusion of 2000 words in Michael West's General Service List. Gardner and Davies reasoned that many of the words on the academic vocabulary list actually appear in Michael West's General Service List.

In general, the work steps of Gardner and Davies NAVL research are as follows. (1) The list of words is determined using lemma. (2) Michael West's General Service List is not excluded from the academic vocabulary list. (3) Corpus data must be large so that in the process of separating high-frequency words, academic words or technical words can be more valid. (4) Academic corpus data must represent the majority of contemporary English. (5) The new vocabulary list must be examined

against academic and non-academic corpus. It aims to determine validity and reliability as a list of academic words.

METHOD

This research uses mixed methods, which is a combination of quantitative and qualitative methods. The quantitative approach usually uses statistical and mathematical test equipment while the qualitative approach is more based on logical reasoning, and understanding of the interpretation of the research object. In other words, mixed-methods involve the analysis of both research methods, namely quantitative and qualitative in a study that seeks to integrate it at one or more stages of the research process (Dornyei, 2007). The source of data is the origin of the data subject obtained (Arikunto, 2016). This study uses research data from two hundred accredited journal articles in Indonesia published by the Decree of the Higher Education which is collected by using cluster sampling techniques in the classification and data mapping. A cluster sampling technique is a technique that considers a group of data to represent the population to be identified and included in the sample (Jackson, 2011). The research data amounted to two hundred scientific articles that are evenly divided into four divisions based on each of the four fields of science namely social, health, engineering, and life science so that there are fifty journals in each field.

Table 2.
Research Data Composition

Knowledge Field	Number of Articles	Word Count
Social	50	239.805
Health	50	128.311
Engineering	50	159.877
Life Science	50	144.794
TOTAL	200	672.787

The research data in the form of journal articles are processed using the Ant Word Profiler corpus program Anthony (2013) which can be downloaded for free. The following are the stages of research data analysis. The first stage, the research focuses on knowing the list of Indonesian academic vocabulary. The second stage, the research aims to make a Indonesia academic vocabulary list. Next stage is a description of the research results.

RESULTS AND DISCUSSION

The Process of Creating an Indonesian Academic Vocabulary List

An academic vocabulary list is a list that contains a number of vocabulary in a specific language used in academic activities (Nation, 2001). In compiling a list of academic vocabulary, linguistic data need to be taken from language activities in the academic world (education) so that the data can represent that field. Academic vocabulary drafting uses the corpus linguistic method. Corpus linguistics is a linguistic research method that works by processing linguistic data in the form of the text so that linguistic data becomes easy to read and analyze (Lindquist, 2018). Text data were chosen because it is easily processed and analyzed because it does not need to be transcribed again. Data processing in corpus linguistics utilizes computer technology so that the data will be easily read in the form of lists and statistics (McEnery, 2019).

The data used in making academic vocabulary lists are linguistic data taken from scientific articles from accredited national journals. A scientific article is a research report text that can be used to represent the academic world. A good scientific article is an article contained in a good journal. To be published in the journal, scientific articles must pass through a selection system and through editing by an editor. Then, the quality of a good journal can be known from the value of its accreditation. A good scientific article will reflect the use of language that is good and right so that the validity of linguistic data can be maintained. Furthermore, national journals were chosen since the authors of articles in national journals were

Indonesians as Indonesian native speakers.

These scientific articles are downloaded from the official pages of each journal. The scientific article is in the form of PDF data so that it can be processed by a corpus machine. The data must be converted into the TXT form using the Ant File Converter application (Anthony, 2013). After all the data is in the form of TXT, the data is cleaned of linguistic elements that are not needed as data, such as the name of the author, journal name, department, faculty, university, logo, email address, etc. In other words, the data used is purely a form of the scientific article so that all forms of identity must be removed. After all data is cleared, all TXT files are merged based on the field of science using the TXT Collector program. This is intended to summarize all the data into a file so that it is easy to process and avoid the risk of errors when processing it into the corpus program. At this stage, the research data are ready to be processed.

Linguistic data processing utilizes the Ant Word Profiler program. Ant Word Profiler is a corpus program made by Anthony Laurence in 2013. The function of the program used in this study is to process linguistic data in the form of text and present statistics and lists. Following is the display of Ant Word Profiler.
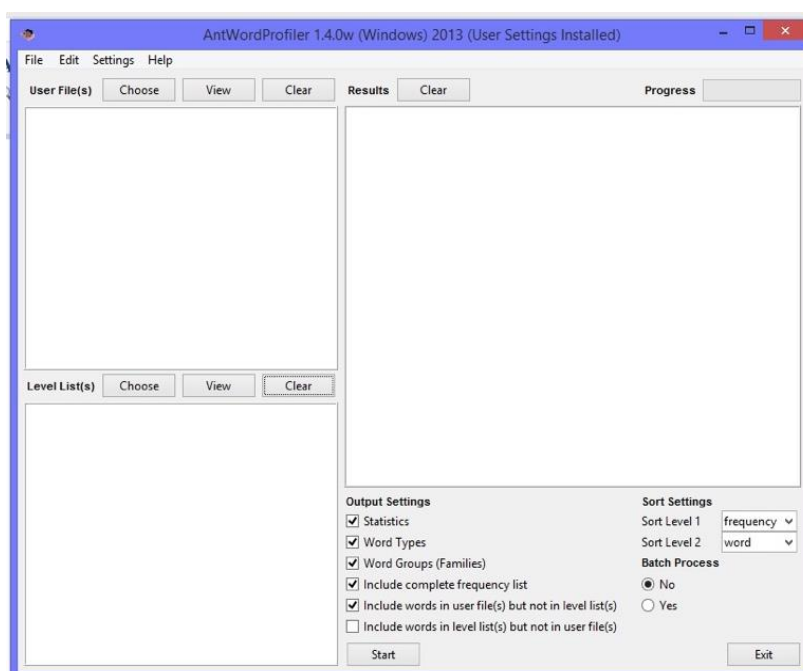

Figure 1. Display of the Ant Word Profiler Program

To make an academic vocabulary list, high-frequency vocabulary is required (Nation, 2001). The Indonesian high-frequency vocabulary was created by Kwary in 2013 under the name Indonesian High Frequency Word List which is presented at KOLITA XI (Kwary, 2013). Indonesian High Frequency Word List is the most frequently used vocabulary list in Indonesian, which means that if someone wants to learn Indonesian, he must master the vocabulary on the list. Indonesian High Frequency Word List is needed to filter the word list in the research data, which means a list of vocabulary used in the academic field outside the Indonesian High Frequency Word List or can be called the Indonesian General Vocabulary List. Next is the data processing using Ant Word Profiler.
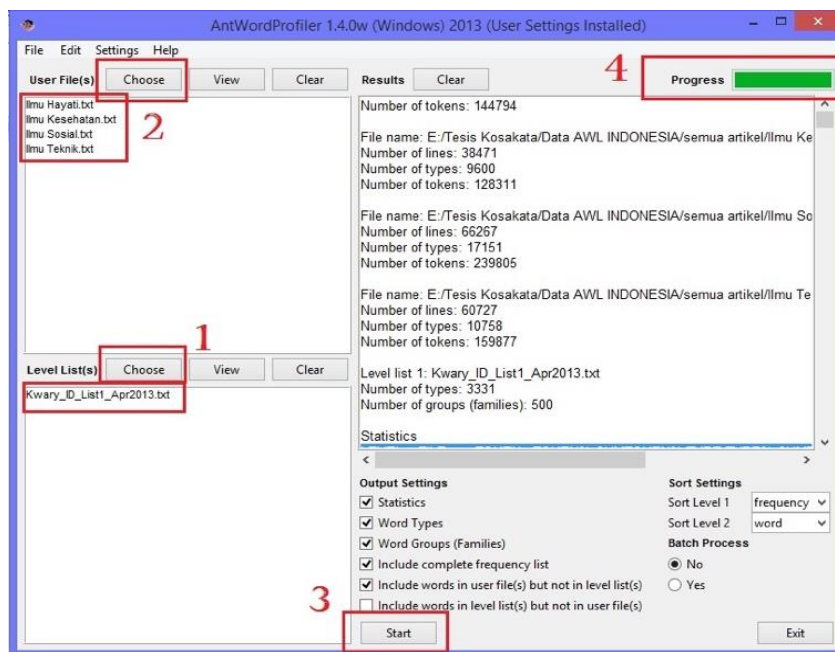
Figure 2. Data Processing with The Ant Word Profiler Program

1. Open the Ant Word Profiler program. Then, in the "Level List(s)" section, click "Choose" button and select the Indonesian High Frequency Word List file (Kwary, 2013).
2. Next, switch to the "User File(s)", click "Choose" button and select all the research data that has been put together based on its sub-exports.
3. Then, click "Start" button.
4. Data processing is in progress. Allow a few moment until the "Progress" indicator turns green which indicates that the processing is complete. Then, the data are ready to read in the "Results" section.

In the results of the data processing, some information will be displayed. There are two things needed in this study, namely statistical information on research data and word lists from research data that are not contained in the Indonesian High Frequency Word List. The list is the forerunner of the Indonesian academic vocabulary list.
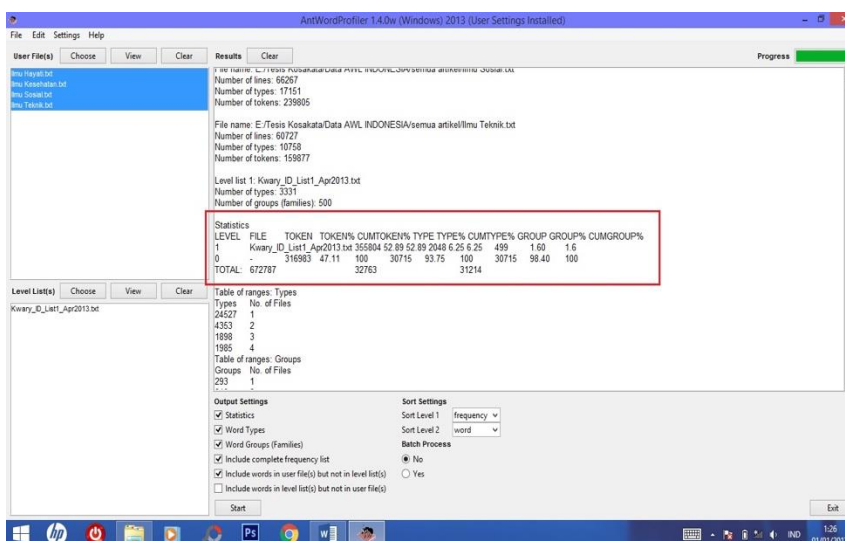

Figure 3. Statistical Data in the Ant Word Profiler

`The statistical results of research data using the Ant Word Profiler program are presented irregularly. This can be overcome by copying text and pasting it into an Excel program. Following is the display after being copied to Excel.

Figure 4. Statistical Data After Moved to Excel

Once copied to Excel, the statistics display becomes organized and easy to read. This statistical data are needed in examining the Indonesian academic vocabulary list.

Other data processing results needed in this study are a list of words that are not contained in the Indonesian High Frequency Word List which is the forerunner to the list of Indonesian academic vocabulary. Due to many data processing results, the referred data search cannot be done manually. First, change the results of the Ant Word Profiler program data processing into a TXT file by clicking the "File" button then selecting "Save Results in Nation 'Range' format ..." then name the TXT file that will be created.
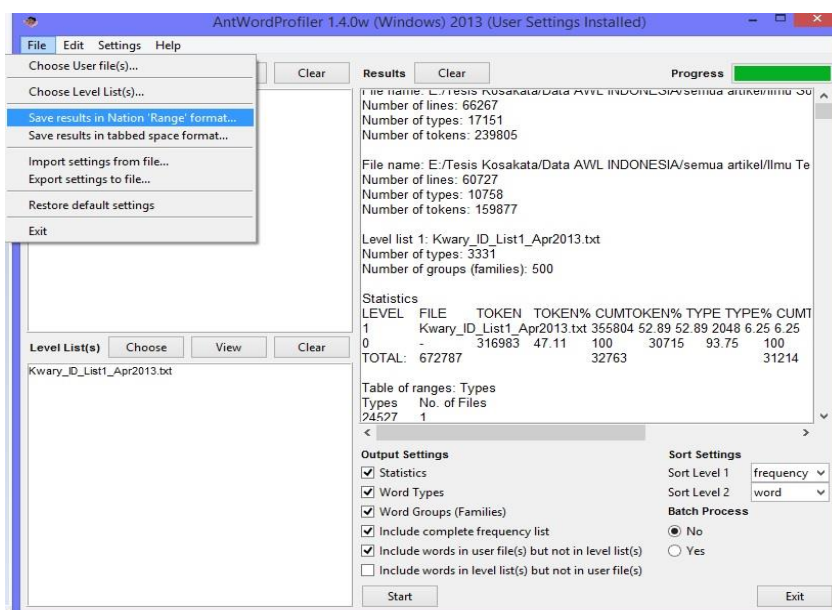

Figure 5. Changing the Format of Ant Word Profiler Results to TXT File

After that, open the TXT file containing the results of the research data processing that was just made. Then, press "ctrl + F button" on the computer. In the search column (Find), type the word "NOT FOUND" then click "Find Next" button as shown below.
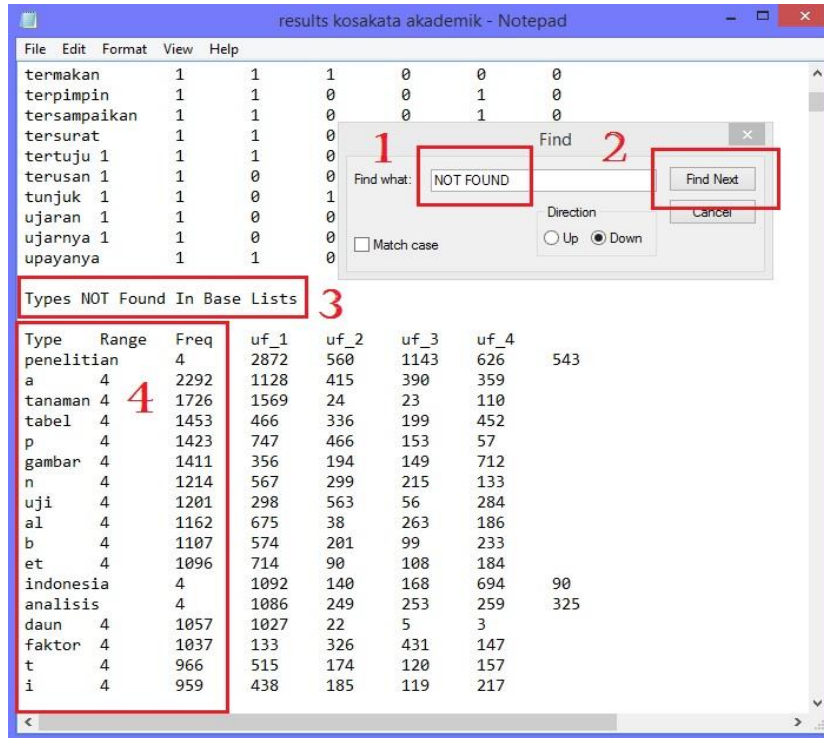
Figure 6. Searching for "Not Found" Data in Notepad (TXT file)

A list of words in "Types Not Found In Base Lists" is a list of words in the text of journal articles that are not in the Indonesian High Frequency Word List. This data is needed in making Indonesian academic vocabulary. The data used in making academic vocabulary lists are listed in the "Type, Range, and Freq" column. "Type" is a list of words in the text of scientific articles that are not in the Indonesian High Frequency Word List. "Range" is the scope of the appearance of words in each field of study (sub-corpora). Frequency is the number of occurrences of words in all scientific article texts. All data contained in the "Types Not Found In Base Lists" are then copied and pasted into the Excel program as shown in the following image.
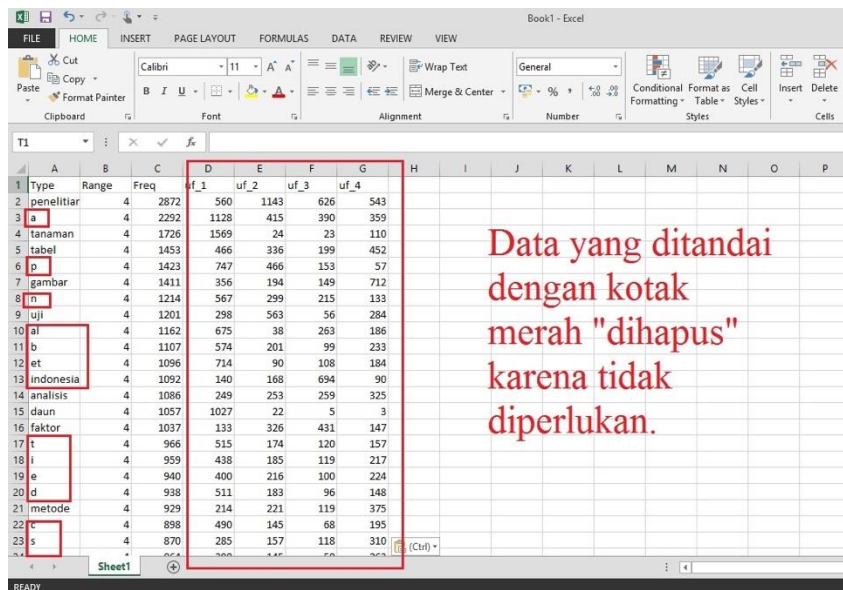

Figure 7. Erasing Unnecessary Data in Excel

After the data processing results are displayed as shown above, the next step is cleaning up data from unnecessary elements such as letters, names (name of a place, people's names, etc), non-standard words, words in other language, abbreviations, etc. After clearing the data from elements that are not needed, the next step is sort the data according to range and frequency from highest to lowest by creating a filter on the column headings "Type, Range, Freq" by blocking the column headings "Type, Range, Freq". Then, click "Home" button, click "Sort & Filter" button, then select "Filter".
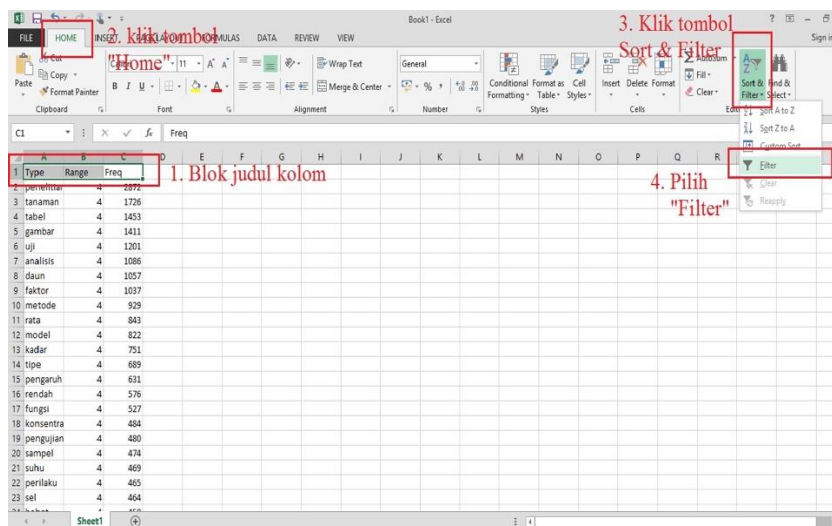

Figure 8. Giving a "Filter" to the Column headings in Excel

The results of research data processing must be sorted by their range from the largest to the smallest. After the range is sorted, the frequencies are also sorted from the largest to the smallest. This ordering should not be reversed because the first criteria from the data compiled into an academic vocabulary list are the words whose range of occurrence in all fields of science (sub-corpora) is the highest, number four representing four fields of science, then down to three, two, and one. After the range is sorted, the frequency is also sorted. The ordering method is as follows.
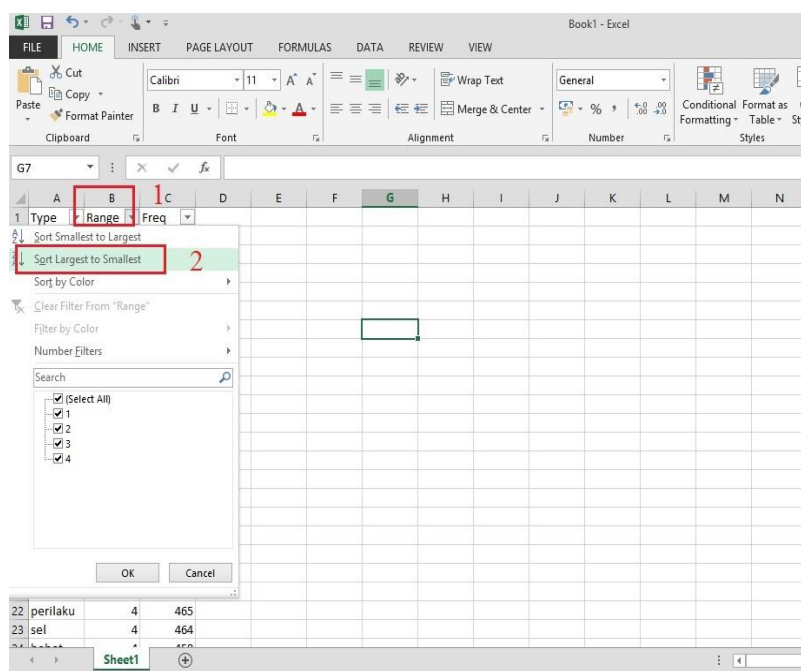

Figure 9. How to sort data "Type", "Range", and "Freq" in Excel

In the "Range" column, click the down arrow button, then select "Sort Largest to Smallest". Do

the same in the "Freq" column.

The next step is the stage of arranging basic words and their derivative words. A list of academic vocabulary is made by taking the list of 250 words from the top list. In compiling Indonesian academic vocabulary, we need to calculate that the number of words in Indonesian is smaller than those in English. Therefore, the number of words on the Indonesian academic vocabulary list does not need to be as much as the academic vocabulary list of Coxhead or Davies and Gardner. The New International Dictionary Webster third edition contains around 472,000 word entries, (Fromkin et al., 2011) the Oxford English Dictionary consists of around 291,500 word entries (Oxford English Dictionary, 2022), while the Large Language Dictionary Indonesia, the fourth edition, only has around 90,000 word entries. Besides, the Indonesian academic vocabulary list is set at 250 words since it refers to the Coxhead academic vocabulary list, which amounts to 570 family words, which is 50% of the total general English vocabulary (High Frequency Word List) made by Nation. In Indonesian, whose vocabulary is less than English, Kwary created an Indonesian High Frequency Word List consisting of 500 family words so that researchers set an Indonesian academic vocabulary of 250 family words, which is a half number of family words in Indonesian High Frequency Word List (Kwary, 2013).

Taking the top 250 words from the academic vocabulary list there should be no words that are still in the same family words. If there are words that are still in the same family words, the word is inserted into the word members whose order is higher.

Figure 10. Determination of Word Family Members

Results of Creating Indonesian Academic Vocabulary

Creating a list of Indonesian academic vocabulary aims to facilitate students, especially BIPA students in writing academic texts in Indonesian. With the Indonesian academic vocabulary list, BIPA students who want to learn in writing academic texts in Indonesian can focus more on what vocabulary they need to master. By mastering 500 word families and their derivatives from Indonesian common vocabulary and 235 word families as well as their derivatives from academic vocabulary, then someone is able to produce academic texts in Indonesian.

From the process of creating Indonesian academic vocabulary as explained in the previous chapter,

we found 235 words which are the basic form of words. A total of 235 Indonesian academic vocabularies are as follows.

1. TELITI	64. TEKNIK	113. TATA	175. SIKLUS
2. TANAM	65. KINERJA	114. CIRI	176. FREKUENSI
3. TABEL	66. MUKA	115. UNSUR	177. BASIS
4. GAMBAR	67. KUNCI	116. KOEFISIEN	178. IDENTITAS
5. UJI	68. KOMUNIKASI	117. BATAS	179. PELIHARA
6. ANALISIS	69. LAJU	118. JARAK	180. GEJALA
7. DAUN	70. NOMOR	119. RELATIF	181. DAMPAK
8. FAKTOR	71. RISIKO	120. DIMENSI	182. ISTILAH
9. METODE	58. LIPUT	121. TEKAN	183. TRANSFORMASI
10. RATA	59. STRATEGIS	122. KOMPONEN	184. RANCANG
11. MODEL	60. TEKNOLOGI	123. MAKSIMAL	185. IDENTIFIKASI
12. KADAR	61. FISIK	124. STATUS	186. PAHAM
13. TIPE	62. PERIODE	125. UNIT	187. OPTIMAL
14. PENGARUH	63. DERAJAT	126. PRIMER	188. BAKU
15. RENDAH	64. STANDAR	127. LAPANG	189. SEDIA
16. FUNGSI	65. INDEKS	128. POHON	190. GRAFIK
17. KONSENTRASI	66. RESPONDEN	129. REKOMENDASI	191. ALTERNATIF
18. SAMPEL	67. CELAH	130. KATEGORI	192. KIMIA
19. SUHU	68. PARAMETER	131. PASANG	193. ORIENTASI
20. PERILAKU	69. ANGKA	132. PERANGKAT	194. BALIK
21. SEL	70. FASE	133. UJUNG	195. HAMBAT
22. BOBOT	71. FORMULA	134. VARIASI	196. VOLUME
23. KONTROL	72. SUKU	135. AREA	197. PROVINSI
24. TOTAL	73. KULIT	136. ENERGI	198. SALUR
25. UMUR	74. PERAN	137. INDIKASI	199. GULA
26. LAHAN	75. TARAF	138. STATISTIK	200. KISAR
27. KONSEP	76. KOLOM	139. ABSTRAK	201. KAJI
28. POLA	77. MUKIM	140. DISTRIBUSI	202. CURAH
29. BATANG	78. KANDUNG	141. SIMPULAN	203. ORGANIK
30. PRINSIP	79. KAPASITAS	142. SKALA	204. SYARAT
31. GANGGU	80. POTENSI	143. CONTOH	205. ALIR
32. KRITERIA	81. TAHAP	144. LABORATORIUM	206. DINDING
33. UKUR	82. EFEK	145. LEVEL	207. PLASTIK
34. MESIN	83. PANAS	146. CAMPUR	208. MEDIUM
35. WARNA	84. GELOMBANG	147. INTENSITAS	209. SOLUSI
36. AMAT	85. NEGATIF	148. KEDELAI	210. BUTIR
37. SKOR	86. RESPON	149. MIKRO	211. MISKIN
38. RUSAK	87. INTERAKSI	150. DEVIASI	212. REAKSI
39. STRUKTUR	88. BAHAS	151. LATERAL	213. UDARA
40. INDIVIDU	89. HITUNG	152. PUTIH	214. HURUF
41. AKAR	90. MEKANISME	153. MATERIAL	215. KUNING
42. OBAT	91. TITIK	154. GARIS	216. EVALUASI
43. SIGNIFIKAN	92. RASIO	155. LEBAR	217. ISOLASI
44. NORMAL	93. MERAH	156. OBJEK	218. LENGKAP
45. POPULASI	94. CENDERUNG	157. KENDALI	219. GOLONG
46. VARIABEL	95. HIJAU	158. TES	220. KONTAK
47. STUDI	96. KORELASI	159. KOMBINASI	221. HAYATI
48. APLIKASI	97. PANEN	160. TIMBUL	222. ENAM
49. ASPEK	98. DEWASA	161. KELOLA	223. FENOMENA
50. JURNAL	99. KOMPLEKS	162. IKLIM	224. IALAH
51. LEMAH	100. PUNCAK	163. KEBUN	225. UNTUNG
52. GLOBAL	101. ALAMI	164. MATRIKS	226. LURUS
53. SARANA	102. KUALITATIF	165. PROSEDUR	227. RESPONS
54. SUDUT	103. LUNAK	167. RIWAYAT	228. ULANG
55. INDUKSI	104. OLAH	166. LETAK	229. BIAS
56. KIRI	105. SEDERHANA	167. TERAP	230. DIAGRAM

57. KULTUR	106. SAJI	168. IRING	231. KEMAS
58. MUTU	107. BASAH	169. SPESIFIK	232. KENDALA
59. AHLI	108. REGRESI	170. EKSPRESI	233. PANGAN
60. PETA	109. MANAJEMEN	171. MINIM	234. PERSEPSI
61. SIMULASI	110. TRANSPORTASI	172. PECAH	235. RINGAN
62. TANGGA	111. JUMPA	173. TAHAN	
63. JANTUNG	112. EFISIEN	174. PENDEK	

The word lists on the Indonesian academic vocabulary list is a specific diction used in the academic field. For instance, the word "total" is more academic than the word "jumlah" which is not in the vocabulary list above. Another example, the academic text uses the word "skor" rather than the word "nilai", "evaluasi" rather than "*penilaian*", "area" rather than "wilayah" or "daerah", and others. Such vocabularies according to Moeliono et al., (2017) are called scientific words and according to Coxhead (2000) and Gardner & Davies (2014) are called academic words.

The results of testing the Indonesian academic vocabulary list in this study are not much different from the results of testing the Academic Word List conducted by Coxhead (2000) in his research. Tokens (words) in the Indonesian academic vocabulary list are 10.8%, exceeding Coxhead's Academic Word List which is only 10%. The test results are also higher than those stated by Nation (2001), who believes that usually, academic words make up almost 9% of the words contained in academic texts. This means that the list of Indonesian academic vocabulary in this study meets the requirements. In other words, by mastering 250 Indonesian academic vocabulary and 500 general Indonesian vocabularies, someone can understand 63.69% of the words in academic texts. Thus, the number of 250 words (not 570 words as in the Coxhead English academic vocabulary list) in the Indonesian academic vocabulary list meets the requirement of 10% of the total words in academic texts.

CONCLUSION

Based on the results of the study, it was concluded that the preparation of a list of Indonesian academic vocabulary can be done using the Ant Word Profiler program. The use of the Ant Word Profiler program facilitates the processing of research data in the form of a large number of vocabulary. In this study, 250 academic vocabularies were produced. The list is a list of the 250 most frequently occurring vocabularies in academic texts. This research can still be developed again by adding the amount of data and variations of the data used, for example adding other types of academic texts.

ACKNOWLEDGMENTS

REFERENCE

Anthony, L. (2013). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Arikunto, S. 2016. *Metodologi Penelitian*. Yogyakarta: Bina Aksara.

Astuti, A. M., & Aziez, F. (2021). Penyusunan Daftar Kosakata Baca Bahasa Indonesia untuk Siswa Sekolah Menengah Pertama (SMP). *Metafora: Jurnal Pembelajaran Bahasa dan Sastra*, *8*(2), 153-164. 10.30595/mtf.v8i2.12415

Browne, C. (2014). A New General Service List: The Better Mousetrap We've Been Looking for. *Vocabulary Learning and Instruction*, *3*(2), 1-10. http://dx.doi.org/10.7820/vli.v01.1.2187-2759

Cindana, S. (2022). Analisis Penggunaan Kosakata Bahasa Indonesia di Kalangan Generasi Alpha. *Ability: Journal of Education and Social Analysis, 3*(3), 43-54. https://doi.org/10.51178/jesa.v3i3.680

Coxhead, A. (2000). A New Academic Word List. *TESOL quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Darwis, M., & Kamsinah, K. (2019). Sumbangan Bahasa Indonesia terhadap Pemekaran Kosakata Bahasa Bugis: Kasus Penggunaan Bahasa Bugis pada Media Sosial Facebook. *SAWERIGADING*, *25*(2), 107-118. https://doi.org/10.26499/sawer.v25i2.626

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

Fromkin, V., Rodman, R., & Hyams, N. (2018). *An Introduction to Language (w/MLA9E Updates)*. Cengage Learning.

Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied linguistics*, *35*(3), 305-327. https://doi.org/10.1093/applin/amt015

Henilia, H. (2023). Pengaruh Kosakata Terhadap Kemampuan Berbicara Bahasa Indonesia. *Madani: Jurnal Ilmiah Multidisiplin*, *1*(7), 532-536.

Herpindo, H., Wijayanti, A., Shalima, I., & Ngestrini, R. (2022). Kategori, fungsi, dan peran sintaksis bahasa Indonesia dengan PoS Tagging berbasis rule dan probability. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *8*(1), 51–65. https://doi.org/10.22219/kembara.v8i1.18602

Indhiarti, T. R., & Chaerunnisa, E. R. (2021). A Corpus-Driven Collocation Analysis of Degree Adverb Very, Really, Quiet, And Pretty. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *6*(2), 226–242. https://doi.org/10.22219/kembara.v6i2.13526

Jackson, S. L. (2011). Research Methods and Statistics: A Critical Approach. *Cengage Learning, 17*.

Kwary, D. A. (2013). Creating and testing the Indonesian high frequency word list. *Jurnal KOLITA XI*.

Kwary, D. A. (2013). The Indonesian High Frequency Word List (Version April 2013). Retrieved from http://www.kwary.net/iwl

Lei, L., & Liu, D. (2016). A New Medical Academic Word List: A Corpus-Based Study with Enhanced Methodology. *Journal of English for academic purposes*, *22*, 42-53. https://doi.org/10.1016/j.jeap.2016.01.008

Lindquist, H. (2018). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

McEnery, T. (2019). *Corpus Linguistics*. Edinburgh University Press.

Mujianto, G., & Sudjalil, S. (2021). Tipe modifikasi fonem kata serapan asing ke dalam bahasa Indonesia pada surat kabar online berbahasa Indonesia. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *7*(1), 1–19. https://doi.org/10.22219/kembara.v7i1.16206

Moeliono, A. M., Lapoliwa, H., Alwi, H., & Sasangka, S. S. T. W. (2017). Tata Bahasa Baku Bahasa Indonesia. Jakarta: Balai Pustaka.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. UK: Cambridge University Press.

Oxford English Dictionary [online]. 2022. www.oed.com, diakses 1 Juli 2022, pukul 08.00 WIB.

Putri, J. J. D., Djatmika, D., & Nugroho, M. (2022). Penanda modifikasi internal pada tuturan encouraging juri Indonesian Idol Special Season dalam memberikan komentar penampilan peserta. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *8*(1), 66–80. https://doi.org/10.22219/kembara.v8i1.18523

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Sari, M. U. K., Kasiyun, S., Ghufron, S., & Sunanto, S. (2021). Upaya Meningkatkan Penguasaan Kosakata Bahasa Indonesia dengan Menggunakan Permainan Anagram di Sekolah Dasar. *Jurnal Basicedu*, *5*(5), 3614-3624. https://doi.org/10.31004/basicedu.v5i5.1425

Selviana, Y., Mannahali, M., & Dalle, A. (2020). Hubungan Antara Penguasaan Kosakata dengan Kemampuan Menyimak Bahasa Jerman. *INTERFERENCE Journal of Language, Literature, and Linguistics*, *1*(2), 148-152.

Supriadin, S. (2016). Identifikasi Penggunaan Kosakata Baku dalam Wacana Bahasa Indonesia pada Siswa Kelas VII di SMP Negeri 1 Wera Kabupaten Bima Tahun Pelajaran 2013/2014. *Jurnal Ilmiah Mandala Education*, *2*(2), 150-161. http://dx.doi.org/10.58258/jime.v2i2.91

Susanto, Y. D. T. (2017). Hubungan Penguasaan Kosakata dengan Keterampilan Berbicara dan Menulis Siswa Kelas IV SDN Gugus Pangeran Diponegoro Kecamatan Ngaliyan. *Universitas Negeri Semarang*.

Sutami, H. (2014). Kamus Besar Bahasa Indonesia Pusat Bahasa; Edisi Keempat. *Wacana, Journal of the Humanities of Indonesia*, *11*(2).

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology with Examples. *Journal of mixed methods research*, *1*(1), 77-100. https://doi.org/10.1177/1558689806292430

Thornburry, S. (2002). *How to Teach Vocabulary*. Pearsons Education Limited.

Valipouri, L., & Nassaji, H. (2013). A Corpus-Based Study of Academic Vocabulary in Chemistry Research Articles. *Journal of English for Academic Purposes*, *12*(4), 248-263. https://doi.org/10.1016/j.jeap.2013.07.001

Webster's Third New International Dictionary (2002). Massachusetts, USA: Merriam-Webster's Inc.

Winarti, S. (2023). Penguasaan Kosakata Bahasa Indonesia Siswa SD Ditinjau dari Aspek Kelas Kata: Studi Kasus pada Tiga Sekolah Dasar di Kota Ternate, Provinsi Maluku Utara. *Tematik: Jurnal Penelitian Pendidikan Dasar*, *2*(1), 6-16. https://doi.org/10.57251/tem.v2i1.890

Yang, M. N. (2015). A Nursing Academic Word List. *English for specific purposes*, *37*, 27-38. https://doi.org/10.1016/j.esp.2014.05.003

Yusuf, I. (2021). Capaian Keberhasilan BIPA Tingkatkan Fungsi Bahasa Indonesia di Kancah Dunia.