ATLANTIS
PRESS

Using Corpus-Based Analysis to Explore the Academic Word Lists (AWL) Emerged in English Manuscript of SBMPTN Tests

Windy Ardianita Sari, Universitas Airlangga

Abstract

The purpose of this study is to investigate the Academic Word Lists (AWL) as a mode in language and learning. The data source of this study were the English manuscript of SBMPTN tests for the last five years (2013 – 2017). Then, the data were analyzed by using AntWord Profiler 1.4.0w to find out the Academic Word Lists and using AntConc.exe to figure out the concordance of the Academic Word Lists appeared, respectively. The finding of this study presented the percent token of Academic Word Lists in each year. Furthermore, this study also found the words which are frequently existed in English Manuscript of SBMPTN tests for the last five years. By recognizing the Academic Word Lists appeared, this study is expected to assist graduated students who prepare their English comprehension in order to get the best result in SBMPTN test.

Keywords: Academic Word Lists; corpus; SBMPTN

1. Introduction

The competition in educational field is increasingly strict. Every student tries to compete each other in order to be able to get in to best school or college as they dreamt about, especially for grade twelve of senior high school students. They only have time which is less than a year to prepare all the materials needed for competing with their peers in the SBMPTN test. Cited from *www.sbmtpn.ac.id*, SBMPTN (*Seleksi Bersama Masuk Perguruan Tinggi Negeri*) is a selection of new admissions through a written exam or combination of written and skill exam result. The implementation of SBMPTN 2018 consists of Printed Writing Exam (UTBC) and Computer Based Writing Exam (UTBK) which can be followed by graduated students in 2016, 2017, 2018 from secondary school (SMA/MA/SMK) and graduated Package C in 2016, 2017, 2018.

One of the material tested for English subject in SBMPTN test is reading comprehension. According to Hu and Nation (2000), comprehensive reading is one of the key of successful language learning. SBMPTN test takers are expected to possess a critical mass of English knowledge, including inferring the meanings of the unknown word from context, identifying the argument structure, and distinguishing idiomatic constructions. In order to comprehend the reading passage well, each test takers must understand at least 80% from the reading passage. Most of learners have found difficulties in apprehending the English words which they rarely hear in everyday use. These words included into Academic Word List (AWL).

The Academic Word List (AWL) was released in the year 2000 by Averil Coxhead from the school of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. The list contains 570 word families which were selected by analyzing a corpus of millions of words from over 400 academic texts. The AWL was developed to provide learners of academic English to target their vocabulary development for all academic disciplines. It was primarily made so that it could be used by teacher as part of a program preparing learners for tertiary level study or used by students working alone to learn the words most needed to study at tertiary institutions.

More research seems to have been conducted on vocabularies and Academic Word List. The research entitled *A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs)* had been conducted by Chen and Ge in 2007. They analyzed a lexical study on the word frequency and the text coverage of the 570 word families from Coxhead's Academic Word List (AWL) in medical research articles (RAs) based on a corpus of 50 medical RAs written in English with 190425 running words. From this

analysis, they found that the text coverage of the AWL words interpreted around 10.07% in English medical RAs and that the academic words used in English medical RAs, that 292 (51.2%) out of the 570 AWL word families were frequently used in English medical RAs and that academic words used in English medical RAs distributed separately throughout a whole RA.

Martinez, et al. (2009) with their study entitled *Academic Vocabulary in agriculture research articles: A corpus-based study* tried to identify the academic words in a corpus of research articles of agriculture. They used Coxhead's (2000) Academic Word List as their guidance to analyze the data. Their result shows that some of the words not only had specific meanings and behaviors related to the genre and also probably the field, but also that some words from the Academic Word List had technical meaning in their corpus. The finding suggests that the need to produce field-specific academic word lists, which is should incorporate all frequent academic lexical items necessary for the expression of the rhetoric of the specific research data. The other research about Academic Word List was conducted by Khani and Tazik. Their research entitled *Toward the Development of an Academic Word List for Applied Linguistics Research Articles* was done in 2013. This research was tried to establish an academic wordlist specific for the field of Applied Linguistics. This research found 773 academic word types. A total of 573 (74.12%) academic words found in the corpus overlapped with the words in Coxhead's AWL (2000).

From the previous studies done with Academic Word List, it seems that there is no study which investigated the Academic Word List in SBMPTN text. In fact, this study is essential to be done for assisting the English learner particularly students who graduated from senior high school to pass through SBMPTN test to achieve the best result on it.

2. Methodology

The data source used in this study was taken from the English manuscript of SBMPTN test. The researcher only took the data which is in English language only, so if in the English manuscript there was another language emerged, it would be ignored. In collecting data for this study, the researcher used some steps. First, the researcher downloaded the English test manuscript file of SBMPTN from the year 2013 – 2017 in the website *https://www.zenius.net/c/2402/soal-spmb-2006-bahasa-inggris*. The downloaded form was still in the *pdf* format. Consequently, the researcher changed the format from *pdf* into *txt* file by using *Ant File Converter*.

After collecting the data, the researcher analyzed them by using *AntWord Profiler 1.4.0w*. The researcher used *AntWord Profiler 1.4.0w* to find out the statistic result of Academic Word List of English manuscript of SBMPTN test from the last five years (2013 – 2017) as well as sorted it which its frequently occurring is three times or more. The researcher only acquired the test manuscript from 2013 – 2017 because the researcher needed to identify in which year the most difficult English test of SBMPTN for the last five years found. Meanwhile, to find the concordance of the particular Academic Word, the researcher used *AntConc.exe*. Finally, the last step was making a general interpretation related to the findings.

3. Result and Discussion

Academic Words contains words that frequently appear in academic texts regardless of subject areas, but it is not common in non-academic materials. Coxhead (2000) conducted a study using an academic corpus from various subject areas and made the best list of academic vocabulary of 570 words. This list does not include words from the General Service Lists. Words such as *assume*, *establish*, and *major* are included in the list. This is referred to as the Academic Word List (AWL), and it covers about 8.5% to 10% of the vocabulary found in academic texts (Coxhead & Nation, 2001). Hence, English learners especially students are strongly recommended to learn the words in the AWL in addition to those in the GSL if they have to read academic texts such as SBMPTN text.

The finding of data provided by the researcher is the statistic result to identify the academic word list in each year of SBMPTN test for the last five years (2013 – 2017). The statistic result is presented in Table 1 (statistic result of Academic Word List). The table also presents the General Service List (GSL) of English Words which is used for defining the standard list of high frequency words. Not only that, it also endows the Academic Word Lists (AWL) which was proposed by Coxhead (2000). The result is focused on the highlight ones. The number that is highlighted means percent token of Academic Word List (AWL).

Table 1: Statistic Result of Academic Word Lists

LEVEL	FILE	TOKEN	TOKEN%	CUMTOKE	TYPE	TYPE%	CUMTYPES	GROUP	GROUP%	CUMGROUP%
Statistics										
1	1_gsl_1st	1418	75.35	75.35	374	58.16	58.16	279	52.54	52.54
2	2_gsl_2nc	91	4.84	80.19	52	8.09	66.25	47	8.85	61.39
3	3_awl_57	147	7.81	88	99	15.4	81.65	87	16.38	77.77
0	-	226	12.01	100.01	118	18.35	100	118	22.22	99.99
TOTAL:	1882			643			531			
Statistics										
1	1_gsl_1st	1286	72.25	72.25	349	54.53	54.53	272	50.94	50.94
2	2_gsl_2nc	101	5.67	77.92	58	9.06	63.59	50	9.36	60.3
3	3_awl_57	130	7.3	85.22	93	14.53	78.12	72	13.48	73.78
0	-	263	14.78	100	140	21.88	100	140	26.22	100
TOTAL:	1780			640			534			
Statistics										
1	1_gsl_1st	1716	73.21	73.21	386	55.62	55.62	295	52.77	52.77
2	2_gsl_2nc	148	6.31	79.52	71	10.23	65.85	57	10.2	62.97
3	3_awl_57	211	9	88.52	116	16.71	82.56	86	15.38	78.35
0	-	269	11.48	100	121	17.44	100	121	21.65	100
TOTAL:	2344			694			559			
Statistics										
1	1_gsl_1st	1716	73.21	73.21	386	55.62	55.62	295	52.77	52.77
2	2_gsl_2nd	148	6.31	79.52	71	10.23	65.85	57	10.2	62.97
3	3_awl_570	211	9	88.52	116	16.71	82.56	86	15.38	78.35
0	-	269	11.48	100	121	17.44	100	121	21.65	100
TOTAL:	2344			694			559			
Statistics										
1	1_gsl_1st	1680	71.67	71.67	439	59.16	59.16	320	53.16	53.16
2	2_gsl_2nd	173	7.38	79.05	73	9.84	69	61	10.13	63.29
3	3_awl_570	244	10.41	89.46	112	15.09	84.09	103	17.11	80.4
0	-	247	10.54	100	118	15.9	99.99	118	19.6	100
TOTAL:	2344			742			602			

In the English manuscript of SBMPTN test of 2013, the percent token of AWL is 7.81%. It is clearly less than the standard of academic text should be. This percentage is supposed to be corresponded to the fiction text or in the level of reading comprehension for junior high school students. Next, in 2014, the AWL of English manuscript of SBMPTN test reached 7.3%. This percentage is obviously decreased from 2013. In fact, in 2013 itself, the English manuscript of SBMPTN test is reflected as non-academic text. Consequently, the level of text difficulty in 2014 is less difficult than 2013.

In 2015, the percent token of AWL seems to show increasing point from the previous year, which is 7.3% to 9%. The text could be regarded as academic text if the percent of Academic Word List (AWL) reached 9% or more. It means, at this point, the English manuscript of SBMPTN test appeared more academic compared to the two year before. The more academic the text is; the more difficult word list is going to be. On the other hand, the less academic the text is, the less difficult word list will be. In the year of 2016, the statistic result of AWL percent token does not experience decreased or increased. It is exactly the same AWL percent token from the year before, which is 9%. It means, there is no developing academic word happened in that two years. It is also concluded that the difficulty level of English manuscript of SBMPTN test from 2015 until 2016 is equal.

Based on the table, it shown that the English manuscript of SBMPTN test in 2017 has the highest percent token of AWL which is 10.41%. This presents an increase of 1.41% from the previous year that is 9%. This improvement is in line with the level of word difficulty in the English manuscript od SBMPTN test. From the statistic result from 2013-2017, the number of percent token of Academic Word List has decrease and increased. Nevertheless, still, the highest number is in the year of 2017. As a result, the most difficult level of English manuscript of SBMPTN test is in 2017.

After demonstrating the statistic result of percent token of Academic Word List, the researcher tries to present what Academic Word List found in English manuscript of SBMPTN test in each year were. The researcher got the Academic Word List result by sorting the word which has the frequency of appearing that is three times or more. In this study, the researcher explained the highest Academic Word which appeared in each year.

Table 2: List of Academic Words

No.	Type	Range	Freq	Uf_1
1	Physical	1	8	8
2	Computer	1	5	5
3	Paragraph	1	5	5
4	Cultural	1	4	4
5	Definition	1	4	4
List of Academic Word in 2013				

No.	Type	Range	Freq	Uf_1
1	Text	1	14	14
2	Energy	1	6	6
3	Paragraph	1	6	6
4	Process	1	5	5
5	Benefit	1	4	4
List of Academic Word in 2014				

No.	Type	Range	Freq	Uf_1
1	Paragraph	1	9	9
2	Option	1	4	4
3	Adapted	1	3	3
4	Author	1	3	3
5	Consume	1	3	3

List of Academic Word in 2015

No.	Type	Range	Freq	Uf_1
1	Areas	1	21	21
2	Professional	1	10	10
3	Benefits	1	6	6
4	Paragraph	1	6	6
5	Percent	1	6	6

List of Academic Word in 2016

No.	Type	Range	Freq	Uf_1
1	Mental	1	23	23
2	Job	1	16	16
3	Globe	1	14	14
4	Issue	1	12	12
5	Paragraph	1	12	12

List of Academic Word in 2017

From the data collected, the researcher has gotten the top five of academic word list in the SBMPTN tests from 2013 – 2017. The explanation of each word emerged is as follows:

a. 2013: **Physical**

The word *physical* occurred 8 times in 2013. If we see the concordance of the word *physical* in AntConc.exe, we can find that the word *physical*, comes with the word *activities* like in the sentence: *the decreasing trend of children to do physical activities at their will* and in the sentence *advice to parents for their children to have enough physical activities*. It means, the word physical in the English manuscript of SBMPTN test in 2013 talk about physical activities. Therefore, the learners do need to understand the word physical as well as the topic which are discussing about physical activities. If the SBMPTN learners comprehend this topic well, at least they can understand the passage provided and they can answer the questions well.

b. 2014: **Text**

In this case, the word *text* is refers to the passage that is provided by the writer of SBMPTN test. Usually the writer wants to know how well the readers can comprehend the text presented. The question appeared often asked on the organization of the text or the meaning of specific word in the text. It can be seen in the concordance of the word *text* in this example: *Which of the following best describes the organization of the text?* or, the writer wants to

examine the test takers on detail information of the text such as *Which of the following is <u>TRUE</u> according to the text?*. So that, the test takers must understand what the text is talking about.

c. 2015: **Paragraph**

The word *paragraph* which frequently emerged in the SBMPTN test in 2015 discussed about specific paragraph of the reading passage. It means, the writer of the SBMPTN test wanted to focus on the specific paragraph in the text, not the whole text. It can be seen in the sentence *Which of the following best restates the idea of <u>paragraph 2</u>?*. It means that the test takers have to comprehend the paragraph 2 in order to answer the question correctly. In addition, the sentence *The argument in the last line of <u>paragraph 4</u> assumes that* showed that the test takers must imply the argument in the paragraph 4 with another sentence which can represent the argument in the paragraph 4 correctly.

d. 2016: **Areas**

The term *areas* in the SBMPTN test in 2016 mostly talked about the natural areas since the concordance of the word *areas* is *natural*. It was proven by the sentence *Today our remaining natural areas serve important roles in the study of ecology, botany, zoology, geology, and soil science.* and in the sentence *Preservation of natural areas also provides many practical benefits.* In this case, the learners have to comprehend the meaing of the word *natural* and *areas*. It means in this reading passage, the writer of SBMPTN test wanted to present the text of which topic is natural areas. By knowing the word which is discussed in the English manuscript of SBMPTN test, the test takers only need to learn the topic appeared in order to do the examination well.

e. 2017: **Mental**

In the year of 2017, the word which was the highest frequently emerged was the word *mental.* The word *mental* followed the word *health* like in the sentence *their family and school consider mental disorder seriously* and in *The survey found concerns about <u>mental health</u> across the country that have doubled alarmingly since 2011.* In this case, the learners have to possess the meaning of the word *mental*, *health,* and *disorder.* It means, in this passage, the writer wanted to deliver the passage of which topic is mental health. If the learners know well this term, they might be possible to answer the question better.

This study also identifies that the word existed from the year 2013 until 2017 is the word *paragraph*. It is logical because almost all of the question discuss about what the topic of the *paragraph* presented in the reading passage is. That is why the word *paragraph* occurs in every English manuscript of SBMPTN test from 2013 to 2017. The result could be seen in the Table 3 below.

Table 3: The word that existed from 2013 – 2017 in SBMPTN Tests

WORD	2013	2014	2015	2016	2017
Physical	8			5	
Paragraph	5	6	9	6	12
Definition, definitions	4 3				
Author	7		3	4	7
Define Defined	3 3				
Environment	3			3	
Text	3	14			3
Process Processes	5	4			
Benefit benefits		4		6	
Adapt Adapted			3		4
Consume Consuming			3 3		
Topic				3	3

4. Conclusion

This study has been done to figure out the Academic Word List of SBMPTN test in the last five years (2013-2017) by using AntWord Profiler. Meanwhile, to find the concordance of the particular Academic Word, the researcher used AntConc.exe. In this study, the highest percent token of Academic Word Lists is in 2017 which is 10.41%. It is considered as academic word because percent token of academic word list is more than 9%. It suggested that the most difficult English manuscript of SBMPTN test is in 2017. Moreover, the word that existed from the year 2013 – 2017 is *paragraph*. By knowing the word that is often appeared in each text of SBMPTN test, the student will have more knowledge and prediction of which topic they should know and learn. This study also contributed for the teachers who educate the students especially the SBMPTN test takers in order to be focused on the material which will be predicted appeared in the next SBMPTN test.

References

Chen, Q. & Ge, G. 2007. 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), 502-514.

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly*, 34, 213-238.

Coxhead, A. & Nation, P. 2001. *The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), Research Perspectives on English for Academic Purposes. (pp.252-267).* Cambridge: Cambridge University Press.

Hu, M. & Nation, P. 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language,* 8(2), 689-689.

Khani, R. & Tazik, K. 2013. Towards the Development of an Academic Word List for Applied Linguistic Word List for Applied Linguistics Research Articles. *RELC Journal,* 44(2), 209 – 232.

Martinez et al. 2009. Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes,* (28), 183-198.

SBMPTN 2018. Available at www.sbmptn.ac.id. (Accessed 20 March 2018).

Soal SPMB 2006 Bahasa Inggris. Available at https://www.zenius.net/c/2402/soal-spmb-2006-bahasa-inggris. (Accessed 1 April 2018).