

## Academic Vocabulary of EFL Students' Writing: A Corpus-based Study

Malikhatul Lailiyah<sup>1</sup>, Lian Agustina Setyaningsih<sup>2</sup>

D3 Bahasa Inggris, FISIP, Universitas Merdeka Malang<sup>1</sup>  
Prodi Ilmu Komunikasi, FISIP, Universitas Merdeka Malang<sup>2</sup>  
Jl. Terusan Raya Dieng No. 62-64, Malang, 65146, Indonesia  
Email: [lian.agutina@unmer.ac.id](mailto:lian.agutina@unmer.ac.id)<sup>2</sup>

**Abstract** - Academic writing is widely believed to have an objective, impersonal, informational view and has a variety of words. The goal is to make readers lighter and faster to understand the content of academic writing. Thus, this study aims at investigating the most frequently used academic words in an English as a Foreign Language (EFL) classroom setting. The authors analyze a corpus of 24 students' writing assignments on argumentative essays. Since this research applies a corpus-based approach, the researchers use Coxhead's (2000) Academic Word List to identify the frequency of those words that appear in students' writing. Then the data is transcribed using AntWordProfiler software and analyzed as a description statistic data. The subject of the study is the fourth semester of Diploma 3 students who were selected with the purposive sampling method. The finding indicates a low variety of academic word families used with the low frequency in this corpus. This study contributes to the search of how non-native speaker writers use academic vocabulary in their work.

**Keywords:** academic vocabulary, corpus-based, efl writing

### 1. Introduction

The study on the vocabulary used in academic writing of non-native students has been received considerable attention. As a result, vocabulary knowledge is considered as an important indicator to measure the quality of the second language in writing (Csomay & Prades, 2018; Guo et al., 2013), in reading (Moghadam et al., 2012; Qian, 1999), in listening (Stæhr, 2009; Zhang & Graham, 2020), and speaking (Alharthi, 2020; Khan et al., 2018; Lailiyah & Setyaningsih, 2020). However, studies mentioned that no matter how useful a vocabulary is, it should not be taught out of context because vocabulary is not just the sum of individual words (Csomay & Prades, 2018; Farady Marta et al., 2019; Nation, 2001).

Generally speaking, academic vocabulary brings great obstacles to students in the teaching and learning of vocabulary. The reasons are (1) because students are not familiar with academic vocabulary since it appears less frequently than general vocabulary items (Coxhead, 2000), and (2) because the academic words are more specific and sometimes more abstract than conversational language (Sibold, 2011). However, most of the researchers agree that there is a strong association between successful students in academic language and academic achievement (e.g. Barnes et al., 2016; Csomay & Prades, 2018; Townsend et al., 2012). Thus, it is necessary to provide essential instruction for Language learners to develop their academic vocabulary. The key roles in setting this goal are by designing appropriate learning activities and selecting the proper materials.

The Academic Word List (AWL) consists of 570-word families compiled by Coxhead (2000). These words are believed to be essential to master for English learners regardless of their field of study (Csomay & Prades, 2018; Youngblood & Folse, 2017). In addition, AWL consists of various academic words that are not common but often exist in a wide range of academic text from various disciplines.

A previous study on a similar topic is found. Csomay and Prades (2018) investigated the academic vocabulary of non-native students' writing. The result indicates a significant relationship between students' scores and the academic vocabulary used in various essay types. Differently, the present study intended to investigate the academic vocabulary used by non-native speaker writers based on Coxhead's (2000) academic word list. Thus, the following research questions guide the investigation:

- 1) What is the coverage of academic words as in AWL items?
- 2) How do the learners use the academic word in their works?
- 3) What are the most frequent words out of AWL items?

## 2. Method

The research argues that corpus is a useful tool to clarify a collection of languages samples that are taken from a specific discourse community, thus it is necessary to provide a language sample that closely matches the language environment of the target audience (Youngblood & Folse, 2017). This study is a small corpus in analyzing the use of Academic Word List (AWL) proposed by Coxhead (2000) on EFL students' essays. Participants of this study are English as a Foreign Language (EFL) students who took Intermediate Reading class. One of the assignments in that class was in a form of composition on an argumentative essay. Students were assigned to analyze a text and to respond by developing their perspective on the issue. The purpose of this activity is to provide critical analysis so that could increase their critical thinking skills.

The researchers selected 24 students' work as a sample of the study among the 65 population. A corpus with a small size is more suitable for qualitative on the use of a certain selection of vocabulary or for preliminary study (Granger, 1998). Thus, this study analyzed the academic vocabulary as in Coxhead's (2000) AWL items qualitatively.

In analyzing the corpus, AntWordProfiler software (Anthony, 2015) was used. AntWordProfiler is a freeware, multiplatform tool for conducting corpus linguistic research on vocabulary profiling. Since this study focuses on Academic Word list (Coxhead, 2000), thus the only word-list uploaded was AWL. However, the software automatically analyzed the word that does not belong to the AWL items.

## 3. Results and Discussion

### Academic vocabulary found on students' writing

As computed by AntWordProfiler Tools, the corpus consisted of the total of 10,343 tokens, 1,306 types, and 1,250 word-families are found in the students' work. Of the total 10,343 tokens, only 576 are academic words. In addition, there are 187 types from AWL. It means that when someone read all those students' work, he would be exposed 576 times academic words with 187 variations of academic words. The remaining type which is more than 80% of the types is another word out of AWL. Taken together, academic word counted only less than 6% of the total running words. Table 1 synthesizes the types, tokens, and groups of words showed in students' work.

Table 1 The result of AntWordProfiler Tools analysis

LEVEL	FILE	TOKEN (%)	TYPE (%)	GROUP (%)
1	Academic vocabulary	576 (5.57)	187 (14.32)	131 (10.48)
Not in the list	Non- AWL	9767 (94.43)	1119 (85.68)	1119 (89.52)
TOTAL:		10343	1306	1250

As revealed by the median, the academic word that appeared at least 2 times in the corpus is considered as the high-frequency item. Thus, these results revealed that there are 74 high-frequency words (those that are above the median), whereas the rest of the 57 academic words are low-frequency words (those that are below the median). While Table 2 presents the 16 most frequently used words, the complete list of AWL word families is presented in Appendix A.

Table 2 The 16 most frequently AWL items used

No	Group	Frequency	No	Group	Frequency
1	create	31	9	expert	14
2	access	28	10	relevant	14
3	conclude	28	11	evident	13
4	technology	22	12	intelligence	13
5	research	21	13	source	13
6	benefit	16	14	compute	12
7	contrast	15	15	evaluate	12
8	author	14	16	impact	12

A close look at students' individual work revealed that the average students' work consist of 430 tokens and the average number of academic words in each students' draft is 24. Below are the deeper analyses of the AWL in students' works. The AWL items are the underlined word.

Example 1:

In today's era, everything is instantaneous and easy to obtain. Technology is growing rapidly, even in just a few seconds we can immediately find what we are looking for, such as useful information, articles and images through Google. But unfortunately, with this search engine, our ability to think deeply about something has decreased. It also affects the way we work, whether doing school or college assignments or even office work. When we use Google, we only get answers but do not get any meaning or lessons because we are only skimming words and pictures without delving deeper first. It also means that we no longer keep facts in our brains because we turn to Google. International experts have put together a new review showing that increasing internet time limits our ability to focus on one task at a time. This is because when we use the internet via cellphones or laptops, of course there will be lots of notifications that appear which eventually make us lose focus because we are distracted. If this continues to be done it will affect our brain function and cause us to think superficially. It is true that Google has a lot of useful information and makes it easier for us. But it is very wrong to think that this makes us think better and faster, when on the contrary it makes us shallow. Although the internet can make us more creative because there are many ideas that we can develop, unfortunately there are still many of us who imitate other people's ideas or work. This can be said to be negative behaviour that is detrimental to others. So even though Google allows us to get all the information easily, it also makes us lazy to use our brains to think deeper. If you really want to be smarter, then the way is to reduce the use of Google. (Student 3)

As explained before, the assignment was asking students to analyze two different arguments and to develop an argumentative paragraph that showed their side of opinion. From the above work, there are 16 academic words of the total 316 running words. The above example shows not only a low percentage of academic words (only 5.06%) but also a less variety of words that showed his/ her perspective in analyzing the argument. Since the goal of this composition is to develop students' critical thinking skills through an argumentative essay, thus it is very important to utilize AWL relating to that.

In many cases of students' work, they use the academic word, however, they fail to utilize it appropriately in their essays. This shows that it is not enough for students to use academic vocabulary, they must also know how to utilize them properly in their essays based on the context and the purpose of writing. In comparison, Example 2 shows work with a high percentage of academic words but with inappropriate use.

Example 2:

Based on the statements that are presented are discussing about two point of views of the most favorite search engine named Google. The first statement says that the google makes us stupid because when we use our computers and our cellphones all the time, so, we're always distracted. But from the other statement says that google makes us smarter we get information for free and can make everything easier. Secondly, both premises as we known as the devil advocacy or "no" and the angel advocacy or "yes" can be classified as believable arguments. However, only "no" side argument that has a high degree of certainty and it is proven in the line "But 81 percent of experts polled by the Pew Internet Research Project say the opportunities outweigh the distractions". "Yes" argument is not fully credible to be categorized as high degree of certainty because it has a lack in serving related research. It can be said that the premises are still built from general point of view. Thirdly, both premises used in the arguments are able to support the conclusion. Then, the second argument offers a better trust to reasonable person because its whole side has a scientific background that covered their side, such as research by Pew Internet Research Project and philosopher's belief from Socrates. Last but not least, the two sides of arguments are strong, both pro and contra have a strong basis so that it can be said to be true depending on the point of view we see it. So that if we look at these two differences of opinion, we can compare and see the truth of this argument that both are true.

Based on the pros and contrast of the arguments above, my argument is Google does not make us to be stupid. It proven scientifically (e.g. Pew Internet Research Project and philosopher's belief from Socrates). Google becomes a platform that is able to facilitating everyone a wide knowledge and learning resources for free. Not only that, as an evidence, an overwhelming majority of 895 experts surveyed by the Pew Research Center's Internet. Their study reveals a factual positive premise in the utilization of Google search engine. The results shows that 3 out of 4 experts said use of the internet enhances and augments human intelligence and two-thirds said use of the internet has improved reading, writing, and rendering of knowledge. Through Google, every scientist is increasingly relying their academic review on finding scholarly literature. Hence, to conclude my review, I would like to give statement that google does not make us stupid. Therefore, the use of google depends on the individual itself, so it can be positive or negative. Instead, Google is widening our knowledge and it is evidenced by previous research, philosopher belief, and annual research basis. (Student 6)

This work consists of 428 running words with 44 tokens of AWL items or 9.3%. The above example is the highest percent of academic words used among all the students' work. However, this paper only has 27 types of AWL items. For

instance, the word “research” is utilized 8 times or 18% of the total academic word used, followed by the word “expert”, “philosophy”, and “project” with 6.8% of each. Overall, this work had less variety of AWL.

As seen from the above student’s work, there are many misuses of words in the using of academic words. For example, the student writes the AWL item “advocacy”, but in this context, the word “argument” would be more appropriate. Moreover, when students use misuses word or make grammatical mistakes on their writing, it will affect the meaning of the text (Putri & Dewanti, 2014). Closer analysis, Student 6 writes an indefinite article “an” for the uncountable noun “evidence”, which is AWL items. To add more, Student 6 uses the AWL item “facilitating” instead of its base form since it is part of the to-infinitive.

Despite the misuse of words and grammatical mistakes, the second example shows that the student has successfully answered the prompt by utilized analysis words such as “evidence”, “contrast”, and “reveals”. Besides, a convincing argument also utilized transition words to build a logical analysis. In this example, the author used the academic word “conclude” to make conclusion, and the use of “hence” to establish a cause and effect.

### The most frequently used word out from the AWL items

The second research purpose of this study is mainly to investigate the high-frequency word in the corpus of students’ writing assignments. The result indicates that there are 9,767 running words that are not in AWL with 1,119 groups of words. It means when the readers read students’ work, 94.43% of the works are not in the academic word list. Below are some examples of the non-AWL items found in the students’ work.

The definite article “the” is the most frequent word used that is mentioned around 558 times. Besides, the indefinite article “a” and “an” is showed up 139 times (rank 14) and 17 times (rank 100) respectively. Example:

“... the second argument has the logic and high degree of certainty...” (Student 5)

“As a human who is always curious...” (Student 10)

“... we can access a lot of information with the blink of an eye...” (Student 18)

Interestingly, Kirana et al. (2018) in their corpus study also found that the article “the” was the most frequently used word.

In the 2<sup>nd</sup> rank, the coordinating conjunction “and” is used 323 times. The other frequently used coordinating conjunctions are “for” (93 times/ rank 15), “but” (55 times/ rank 32), “so” (53 times/ rank 34), “or” (52 times/ rank 36). On the other hand, the coordinating conjunction “nor” is only found 1 time and no coordinating conjunction “yet” is found. Example:

“...the first argument is strong enough for those who do not think deeply but just a glimpse.” (Student 2)

“Even some people said that googling can distract your brain and make you not think deeply because...” (Student 4)

## 4. Conclusion

This corpus study investigated the academic vocabulary use in EFL students’ writing. The Academic Word List by Coxhead (2000) was adopted to calculate the use of academic vocabulary on students’ assignments. Overall, the academic vocabulary used by the students was 5.57% of the running word. Meanwhile, a closer look at students’ individual work showed that most of the students’ works were not only had a lower percentage, but also less variety of academic words. In some cases, the use of the words is also inappropriate in making sense of the argumentative essays. Needless to say, it is not enough for students to just embellish their work with academic words, but it is also necessary to know how to use them properly according to the context and purpose of their writing.

Since the limitation of this study is on its sample size, it is suggested to have similar research with the large sample size. Finally, this study only examined the academic words use compared to non-academic words in students’ assignments. Further analysis can be carried out on the deeper analysis of the context and the use of words as they relate to the overall quality of the text.

## References

- Anthony, L. (2021). AntWordProfiler (Version 1.5.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Alharthi, T. (2020). Investigating the relationship between vocabulary knowledge and FL speaking performance. *International Journal of English Linguistics*, 10(1), 37–46. <https://doi.org/https://doi.org/10.5539/ijel.v10n1p37>
- Barnes, E. M., Grifenhagen, J. F., & Dickinson, D. K. (2016). Academic Language in Early Childhood Classrooms. *Reading Teacher*, 70(1), 39–48. <https://doi.org/10.1002/trtr.1463>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes*, 33, 100–118. <https://doi.org/10.1016/j.jeap.2018.02.003>
- Farady Marta, R., Hafiar, H., Budi Setiawan, Y., Andriani, F., Lestari, P., Pamungkas, S., Ratri Rahmiaji, L., Alif, M., Yuli Purnama, F., & Agustina Setyaningsih, L. (2019). Author compliance in following open journal system of communication science in Indonesia. *Journal of Physics: Conference Series*, 1175(1), 0–10. <https://doi.org/10.1088/1742-6596/1175/1/012222>

- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: Theory, Analysis and Applications*, 145–160.
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218–238. <https://doi.org/http://dx.doi.org/10.1016/j.asw.2013.05.002>
- Khan, R. M. I., Radzuan, N. R. M., Shahbaz, M., Ibrahim, A. H., & Mustafa, G. (2018). The role of vocabulary knowledge in speaking development of Saudi EFL learners. *Arab World English Journal (AWEJ) Volume*, 9. <https://doi.org/https://dx.doi.org/10.2139/ssrn.3151128>
- Kirana, D. P., Basthomi, Y., Isnawati, U. M., & Fitriani, A. (2018). A Corpus-based Study of Vocabulary as Input in EFL Text-Book : A Case in an Indonesian Islamic College. *Journal of Applied Linguistics and Language Research*, 5(5), 93–104.
- Lailiyah, M., & Setyaningsih, L. A. (2020). Students' perception of online communication language learning through Instagram. *EnJourMe (English Journal of Merdeka) : Culture, Language, and Teaching of English*, 5(2), 188–195. <https://doi.org/10.26905/enjourme.v5i2.5202>
- Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia-Social and Behavioral Sciences*, 66, 555–563.
- Nation, P. (2001). Learning Vocabulary in Another Language. In *English for Specific Purposes*. Cambridge University Press. [https://doi.org/10.1016/s0889-4906\(02\)00014-5](https://doi.org/10.1016/s0889-4906(02)00014-5)
- Putri, P. S., & Dewanti, A. (2014). An analysis of grammatical errors in writing narrative texts done by the second semester students at the Diploma Program English Department in Airlangga University Surabaya. *Journal Anglicist*, 3(01), 1–7.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 282–308.
- Sibold, C. (2011). Building English Language Learners' Academic Vocabulary: Strategies and Tips. *Multicultural Education*, 18(2), 24–28.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577–607. <https://doi.org/https://doi.org/10.1017/S0272263109990039>
- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 112(3), 497–518.
- Youngblood, A. M., & Folse, K. S. (2017). Survey of Corpus-Based Vocabulary Lists for TESOL Classes 1. *MEXTESOL Journal*, 41(1), 1–15.
- Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning*, 70(4), 1017–1053. <https://doi.org/https://doi.org/10.1111/lang.12411>