

The Distribution of Academic Vocabulary List (AVL) in Undergraduate Theses: A Corpus Study

Fisma^{1*}, Abidin Pammu¹, Ria Jubhari¹

¹ Hasanuddin University, Indonesia

*Correspondence: fismaasdira@gmail.com

ABSTRACT

This study is a corpus-based study that aims to find out the use of words in Gardner and Davies' (2014) Academic Vocabulary List (AVL) in Undergraduate Theses in the field of English Education. A 2.2 million-word called undergraduate theses corpus was created in this study. The corpus consists of 200 undergraduate theses compiled from 10 Universities in Indonesia. The result shows that AVL words account for 24.79% of the entire education undergraduate theses, which contained 557,658 running words. AVL contains 3,015 core academic words (lemmas). Based on word frequency criteria, this study found 573-word forms and 871 lemmas, which coverage 31.48% of the whole Gardner and Davies' AVL that occur at least 63 times in the entire corpus and the word has to occur at least five times in five universities. The finding shows that AVL is effective for the students in writing undergraduate theses. However some of the words that are closely related to the English Education field are not found in the AVL, therefore an attempt to develop a new academic vocabulary specific to the English Education field is needed.

ARTICLE HISTORY

Published March 27th 2021



KEYWORDS

Academic Vocabulary, Theses, Corpus Study, non-native learners, academic writing

ARTICLE LICENCE

© 2021 Universitas Hasanuddin
Under the license CC BY-SA
4.0



1. Introduction

Learning English as a foreign language poses a challenging aspect for students and lecturers, particularly in Indonesia (Rahman, 2018). One of the keys to succeeding in understanding the text is sufficient vocabulary. The ability of student academic vocabulary at the university is a crucial role to pay attention because the most predominant language problem of non-native learners in academic writing is determining which words are academic word and the density of unknown words in academic text, The problem of academic vocabulary also came from the teachers. The teacher in university should grasp the need of students in learning academic words. Academic vocabulary is one of the most challenging aspects of making decisions in which words are worth teaching (Coxhead, 2000; Vongpumivitch et al., 2009). The teachers do not know the occurrences of the word and which words are genuinely representative in teaching new vocabulary (Chanasattru & Tangkiengsirisin, 2017). The sources determine the effectiveness of teaching academic vocabulary. To find out the use of academic vocabulary used by language learners, it is important to identify students' academic vocabulary in the resulting written work, such as in the undergraduate thesis.

Academic vocabulary is an essential element in University students because it is used to write academic text, mainly writing theses. Academic vocabulary "refers to a set of lexical items that do not core words but are relatively frequent in academic texts" (Paquot, 2010; Prihandoko, et al., 2019).

It commonly appears in a large number of academic texts, such as journal articles, theses/dissertations, research papers, conference papers, academic books, etc (Rahman, 2019). However, the students tend to be unfamiliar with academic vocabulary because it has low frequency than general vocabulary. Consequently, knowing academic vocabulary is a demanding task for students to be acquired. The students in the undergraduate programs with minimum experience in writing academic texts have some difficulties writing an undergraduate thesis (Rahman et al., 2019). Sometimes the students find them difficult to distinguish which words are academic and which are general words. The preliminary study reported that choosing the appropriate term is one of the challenges in writing theses. Sometimes students also found difficulties in paraphrasing the idea and looking for the word's appropriate synonym to avoid plagiarism. The students also rarely used theses as references for academic Writing (Prihantoro, 2016).

Gardner and Davies (2014) found a new Academic Vocabulary List (AVL) taken from a larger corpus, which contains 120 million academic texts of the 425-million-word Corpus of Contemporary American English (COCA; Davies 2012) which involve general discipline. The corpus is composed of nine disciplines: Education, Humanities, History, Social Science, Philosophy, Religion, Psychology, Law and Political Science, Science and Technology, Medicine and Health, Business, and Finance. All the text in the academic corpus was compiled from the USA. Academic Vocabulary List (AVL) was created by considering four criteria. 1) Ratio: The frequency of the word must occur at least 50% higher in the academic corpus than in the non-academic reference corpus of COCA (per million words). 2) Range: The word must achieve at least 20% of the expected frequency in at least seven out of nine academic disciplines represented. 3) Dispersion: the words must have a dispersion of at least 0.80 to measures that the word occurred evenly in the corpora or superior to the range measures. 4) Discipline Measures: a word could not appear more than three times the expected frequency in any of nine disciplines.

Some scholars had developed a new academic vocabulary in general disciplines (Gardner and Davies, 2014; Coxhead Averil, 2000). However, academic vocabulary compiled from multiple disciplines seems not effective, therefore some scholars also try to develop a new academic vocabulary in a specific discipline, for example, Lei & Liu (2016) develop a new academic word list in the field of medical, It-ngam & Phoocharoensil (2019) in the field of Science, et cetera. Some of the scholars also tried to investigate the distribution of academic words but focused on Coxhead's AWL in several fields such as in Social Science (Chanasattru & Tangkiengsirisin, 2017) and education (Mozaffari and Moini, 2014) Environmental (Liu & Han, 2015); Agricultural Research Articles (Martinez et al., 2009); and financial corpus (Li & Qian, 2010)

The previous research shows that all of the scholars were interested in investigating the distribution of Academic Word List from Coxhead (2000), but only a few scholars tried to investigate the distribution of Academic Vocabulary List from Gardner and Davies (2014) that more update.

Nowadays, only a low substantial of researchers have attempted to involve Gardner and Davies' (2014) AVL in their research. Csomay & Prades (2018) have attempted to investigate academic vocabulary in ESL student paper. They reported that the overall AVL use in ESL student paper was 12.03%. The data have shown that the contribution of AVL in students' papers is highly skewed and useful.

This study is a corpus-based analysis that explores the use of Academic Vocabulary List (AVL) in undergraduate theses from the English Education Study Program that can help the students acquire academic words to enrich their competence in writing an undergraduate thesis. The result of this research can help the teacher to teach the academic words effectively and analyze which words their students used in their academic text. This study reports vocabulary that the students know, so the teachers can analyze what items of vocabulary that the learners need to know as well as can use in Academic Writing subject as material and designing syllabus.

2. Method

The corpora used in this study include a 2.2 million-word corpus composed of 200 undergraduate theses from 10 Universities in Indonesia published between 2016-2020. The undergraduate thesis had Abstract – Introduction - Literature Review - Finding and Discussion – Conclusion and Suggestion. Some parts are excluded such as numbers, table of content, references, cover, and et cetera. The units of analysis were tokens, range, and frequency.

The undergraduate theses in each University were combined in one file. After that, the corpus is inserted into the software. For the study, the computer software used for analysis was LancsBox 5.0 (Brezina et al., 2020) as a tool to know the occurrence of word frequency and range. This is a suite of the program that can cover the need for this analysis. The software showed the occurrences of each word from the corpus. Then, the data was copied into Ms. Excel for easier to be investigated.

The purpose of this study is to find out the distribution of AVL in undergraduate theses. In doing so, the first stage is to obtain a list that met word selection criteria adopted from Coxhead (2000). The frequency criterion is that the word forms have to occur at least 63 times in the entire corpus. The word range refers to the occurrence of the word in each of the sections (or sub-corpora) of the corpus (Nation & Webb, 2011). In the present study, the range criterion is that the word forms have to occur at least five times in five universities. All of the words that are not met word selection criteria removed from the list.

3. Finding and Discussion

Table 1 represents AVL coverage in undergraduate theses accounts for 24.79% of the entire education undergraduate theses, which contained 557,658 running words. AVL contains 3,015 core academic words (lemmas). Based on word frequency criteria, this study found 573-word forms and 871 lemmas, which coverage 31.48% of the whole Gardner and Davies' AVL. The coverage of the AVL in undergraduate theses is higher and the percentage shows that the use of AVL is effective as references in writing undergraduate theses.

Table 1. The distribution of AVL in Undergraduate theses

	Tokens	Percentage	Lemmas	Headwords
Selected Words	1030954	45.82%	1680	1192
AVL	557658	24.79%	871	573

The most frequent AVL in undergraduate theses was research, which occurs 24,191 times in entire universities. It is the same with AWL in undergraduate theses, which covers 4.34% of AVL in the whole undergraduate thesis. Meanwhile, the top ten of the most frequent word that occurs in AVL (except *research* and *data*) differed with the word in AWL. The word of *learn, use, read, language, mean, class, study, and word* are academic vocabulary words that frequently occur in AVL, but they are not included as an academic word in AWL. Meanwhile, these words correlate with the field of education. Table 4 shows the statistical results of the top 10 AVL in undergraduate theses and the full list can be seen in Appendix

Table 2. Top 10 most frequently used AVL in undergraduate theses

Rank	Word Family	Frequency	Range	Coverage of AVL
1	Research	24,191	10	4.34%
2	Learn	21,081	10	3.78%
3	Use	20,395	10	3.69%
4	Read	10,198	10	1.83%
5	Language	9,035	10	1.62%
6	Data	7,730	10	1.39%
7	Mean	7,398	10	1.33%
8	Class	7,310	10	1.31%
9	Study	7,208	10	1.29%
10	Word	6,985	10	1.25%

The following passage is an excerpt selected randomly from an undergraduate thesis. The word of AVL in the undergraduate thesis is in bold. The passage contains 109 tokens, 55 types, and 49 lemmas. Among the total 109 words, 19 words in the passage were included in academic vocabulary in undergraduate theses. The coverage of the AVL in this text is 21.79% of the total AVL in undergraduate theses.

A method to analyze the error is called error analysis. Error analysis is used to identify the error made by the language learner. As a related statement before that error analysis is used to identify what the student's error made. Brown in Danurwinda (2014:2) stated the fact that learners do make errors, and that these errors can be observed, analyzed, and classified to reveal something of the system operating within learner, led to a surge of study of learner errors. It means that the error is not a mistake that cannot be explained. The error can be observed, analyzed, and classified to know what the student's error made.

Gardner and Davies (2014) develop academic vocabulary, which involves large corpora. A total of 120-million words had compiled in their study from various disciplines. It means that this corpus was almost 35 times larger than the 3.5-million word designed by Coxhead' AWL. The finding shows that not all the AVL words closed-related to the Education word in English Education Study Program. Some words have a close correlation with the education field but are not included in the AVL, such as *material, classroom, competence, curriculum, pronunciation, syllabus, and pedagogy*. The finding also found that some words in AVL still contain general words such as the word of *day, name, age, and help*. It can be concluded that providing academic vocabulary from the general field seems ineffective because some words which are frequent in one field may be absent in another field (Xue and Nation, 1984), and the students in different areas have

different needs. Therefore it is very important to develop a new list in the specific field, particularly in the English Education Study Program

4. Conclusion

The present study is concerned with Gardner and Davies' AVL's distribution involving academic vocabulary for undergraduate students in writing a thesis. The result of this study found that the percentage of AVL in the undergraduate thesis is very high. However, some words are very closed-related in English Education Study Program but are not included as a part of academic words in AVL because AVL was built from general disciplines. However, the result of this study can provide a picture academic words in undergraduate theses that be utilized for designing syllabus and as references in teaching and learning academic words in University. As a matter of fact, an effort to develop new academic vocabulary in the specific field is very crucial, therefore the researcher suggested investigating the new academic word which involves a massive number of academic text from the specific field in future research.

References

- Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Chanasattru, S., & Tangkiengsirisin, S. (2017). The Word List Distribution in Social Science Research Articles. *Arab World English Journal*, 8 (4), 412–429. <https://doi.org/10.24093/awej/vol8no4.28>
- Coxhead, Averil. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213. <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>
- Davies, M. 2012. 'Corpus of Contemporary American English (1990–2012),'available at<http://corpus.byu.edu/coca/>. Accessed March 2020.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*,35 (3), 305–327. <https://doi.org/10.1093/applin/amt015>
- It-ngam,T.,& Phoocharoensil, S. (2019). The development of science academic word list. *Indonesian Journal of Applied Linguistics*,8(3),657–667. <https://doi.org/10.17509/ijal.v8i3.15269>
- 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>
- Li,Y.,& Qian,D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, 38 (3),402–411. <https://doi.org/10.1016/j.system.2010.06.015>
- Liu,J., & Han,L. (2015). English for Specific Purposes A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39, 1–11. <https://doi.org/10.1016/j.esp.2015.03.001>
- Martinez,I.A., Beck,S.C., & Panza,C.B. (2009). Academic vocabulary in agriculture research articles : A corpus-based study. *English for Specific Purposes*, 28,183–198. <https://doi.org/10.1016/j.esp.2009.04.003>
- Mozaffari, A., & Moini, R. (2014). Academic Words in Education Research Articles: A Corpus Study. *Procedia-Social and Behavioral Sciences*, 98,1290–1296. <https://doi.org/10.1016/j.sbspro.2014.03.545>
- Nation, P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Paquot,M. (2010). *Academic Vocabulary in Learner Writing*. Research in Corpus and Discourse. London: Continuum International Publishing Group.
- Prihandoko, LA., et al. (2019). English language competence for tourism sector in supporting socio-economic development in Merauke: A Survey Study. IOP Conference Series: Earth and Environmental Science 343 (1), 012170
- Prihantoro. (2016).The Influence of Students' L1 and Spoken English in English Writing: A Corpus-Based Research. *Teflin Journal*, 27 (2). 217-245.
- Rahman, F. (2018). The Constraints of Foreign Learners in Reading English Literary Works: A Case Study at

Hasanuddin University. *Journal of Arts & Humanities (JAH)*, 7(2), 01–12.

Rahman, F. (2019). Meretas Jalan Publikasi Jurnal Internasional bagi PTN-PTS di Kota Watampone. *Jurnal Ilmu Budaya*, 7(1), 146–151.

Rahman, F., Amir, P., & Tammasse. (2019). Trends in Reading Literary Fiction in Print and Cyber Media by Undergraduate Students of Hasanuddin University. *International Journal of Education and Practice*, 7(2), 66–77.

Vongpumivitch, V., Huang, J.yu, & Chang, Y.C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33–41. <https://doi.org/10.1016/j.esp.2008.08.003>