3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 30(4), December 2024 http://doi.org/10.17576/3L-2024-3004-20

Unlocking Academic Vocabulary: Corpus Insights from Open and Distance English Language Learning Coursebooks

HAMISU HAMISU HARUNA
Department of English, College of Humanities
Al-Qalam University Katsina, Nigeria

AZZA JAUHAR AHMAD TAJUDDIN *
Center of Fundamental and Continuing Education
Universiti Malaysia Terengganu, Malaysia
azzajauhar@umt.edu.my

CHAIROZILA MOHD. SHAMSUDIN
School of Liberal Studies (PPCU)
Universiti Kebangsaan Malaysia, Malaysia

ABSTRACT

Finding the right word that is neither too difficult nor too simple for English learners is one of the most serious problems in applied linguistics, given the undeniable association of vocabulary with comprehension, problems about which academic words need to be used in open and distance learning coursebooks for students' effective reading comprehension remain unsolved. This study serves this purpose by conducting data-driven research on the coursebooks of the National Open University of Nigeria. Therefore, this study uses Coxhead's (2000) Academic Word List (AWL) in a corpus of 6,802,300 words from first-year university coursebooks. The results indicate that while the AWL contains 9.47% of the lexical items in the corpus, certain items on the list frequently occur and exhibit distinct behaviours across disciplinary fields. Using frequency and range selection criteria, the study finds 181 AWL and 28 non-AWL word items used at least 1132 times in the NOUNC and at least 15 times across the 30 academic subjects. This study uncovers cross-disciplinary lexical features that can be used to develop vocabulary acquisition and help students participating in open and distance learning courses improve their academic reading and other language skills. In light of the above, the AWLNOUN could serve as a reference resource for educators and course designers in curriculum development for open and distance learning centres.

Keywords: academic vocabulary; academic word list; AntWordProfiler; corpus analysis; English language coursebooks

INTRODUCTION

The acquisition of a broad and varied vocabulary continues to be postulated as an essential prerequisite for learning and mastering a foreign language, a view that has been widely discussed in academic literature for three decades (Cunalata Guilcapi, 2023). For this reason, the acquisition of vocabulary is considered one of the essential elements of language mastery. This has resulted in the development of various vocabulary-learning tools, as shown by Cetinkaya and Sutcu (2019). As English becomes an international language of communication or a language used in all parts of the world (Haruna et al., 2018), vocabulary can be seen as the basis for meaningful discourse, indicating the integration of lexical and communicative knowledge. This symbiotic relationship highlights the crucial barrier that arises when there are few words that are familiar to learners at an academic level, which slows down their mastery of a foreign language (Waluyo & Bakoko,

2021). Therefore, language learners, especially those who have difficulty communicating in a foreign language, need to develop a rich lexical repertoire to overcome the challenges associated with the target language.

In the current academic context, English has established itself as the leading language for professional discussion and publication, maintaining its status as a medium of instruction in universities. This global phenomenon has generated a growing literature on vocabulary knowledge, which is considered a central component of language acquisition in many studies (e.g. Haruna et al., 2024a). The relationship between vocabulary and language proficiency has been extensively demonstrated. It also emphasises that lexical knowledge cannot be dispensed with when it comes to comprehension regardless of the developmental stage of grammatical acquisition of the target language (Haruna et al., 2024a).

Previous studies have shown that knowledge of word meanings has an important function in promoting reading comprehension and helping learners draw new information from academic texts (Panmei, 2023). These empirical findings support the effectiveness of the academic aspect of vocabulary knowledge in predicting academic performance, making it essential for learner acquisition and the acquisition of new knowledge through language proficiency (Haruna et al., 2024b). The development of a rich and extensive store of meaningful words is, therefore, significant for the acquisition of a new language.

In this regard, this study offers important insights as it examines the academic vocabulary in the coursebooks of the National Open University of Nigeria (NOUN) – a university that has pioneered the provision of open and distance education in Nigeria. Therefore, this study seeks to examine the vocabulary used in the context of open and distance education by profiling the academic vocabulary of National Open University of Nigeria coursebooks (NOUNC). The implication of this study is to improve language practices across various educational modalities. The results will help in developing language learning and teaching resources that complement the observed form of instruction. In addition, these ideas will be used in the development of curricula, teaching methods or other pedagogical tools based on the paradigms of this rapidly growing form of education.

Although previous studies have examined academic vocabulary in traditional classrooms, less is known about the same in Open and Distance Learning (ODL) materials. This justifies the need for this study to develop a list of academic vocabulary for ODL. This is because, during their university education, students in ODL centres are required to read academic texts such as coursebooks and write academic papers and articles in English for course assignments and projects. In the present study, analysing the lexical composition of NOUNC will help to understand the specific lexical challenges that learners face in their studies and provide insights into instructional practices and materials design that could help to improve the processes through which learners in ODL universities build essential academic vocabulary. This study is relevant to the field of applied linguistics and to expanding the scope of further research in the area of vocabulary development and enhancement under ODL settings with a focus on elaborating the necessary vocabulary instruction and materials.

Since this study examines academic vocabulary in the ODL context, the basic conceptual framework for academic vocabulary is presented in the next section. Then, in the methodology section, the corpora and the method of data analysis are presented. In the results section, the research findings are presented. This is followed by a discussion and, finally, a section on the limitations and implications.

LITERATURE REVIEW

The centrality of academic vocabulary to present and future academic achievement has been recognised in the current learning environment, which has led to a paradigm shift in two areas: language acquisition and English for Academic Purposes (EAP). With the renewed globalisation of the English language (B. Ibrahim et al., 2018a), the academic word list (AWL) has become an important predictor of academic achievement (Coxhead, 2000). This prominence is attributed to academic vocabulary being identified by I. S. P. Nation (2001) as the most crucial among the four categories of English requisites linked with students' academic pursuits. Consequently, both researchers in the fields of vocabulary and EAP have recently directed their attention towards academic vocabulary. Academic vocabulary refers to a specialised subset of high- and low-frequency vocabulary used in the English language for specific purposes (Coxhead & Nation, 2001). In other words, this type of vocabulary lies between low-frequency and high-frequency words (Chabbi & Boukezzoula, 2023). Warnby (2022) further defines academic vocabulary as words that are frequently used in a variety of disciplines with a wide range of words.

According to Schmitt et al. (2011), it is crucial for research that students at all educational levels understand academic vocabulary. Academic reading comprehension can be enhanced by learners' familiarity with the academic vocabulary used in academic texts (Schmitt et al., 2011), especially for students enrolled in ODL courses. In addition, research shows that possession of academic vocabulary improves students' long-term performance and the quality of their academic writing. Similarly, Townsend et al. (2012) found a correlation between students' low academic performance and their limited academic vocabulary.

In light of the above, since vocabulary is associated with words used by learners of a particular language, academic vocabulary can be seen as the collection of words used for academic writing. It has been emphasised that students are better able to read, write, speak and listen in a disciplinary way when the text they read makes careful use of academic words (Brandenburg-Weeks & Abalkheel, 2021). Academic vocabulary can, therefore, be a good indicator of how well students perform on writing and reading tasks (Bahtiar et al., 2020) as well as their overall academic performance across educational levels (Schuth et al., 2017). Thus, learners are required to comprehend core vocabulary to grasp the academic themes of their studies. However, it has been found that many students in various academic settings are unable to acquire academic vocabulary, especially in ESL contexts (Choo et al., 2017). In addition, a link has been demonstrated between students' writing quality and their use of academic language (Haruna et al., 2024a). Knowledge of academic vocabulary also has a significant impact on academic success (Schuth et al., 2017). Thus, it is considered challenging for language learners as it plays a solid and robust supportive function (Malmstrom et al., 2018). This is true because academic vocabulary terms are less common than general high-frequent words (I. S. P. Nation, 2001).

Given the relevance attached to academic vocabulary for advanced students in general and its subject-specific nature, researchers have developed academic vocabulary lists based on a set of word selection criteria, with Coxhead's (2000) AWL as a prime example, and since then, interest in identifying the most critical and necessary academic words for EAP success at different levels continues (Kamasak & Soyaltin, 2021). Therefore, due to the pressing demand for vocabulary in higher education ODL contexts, the present study aims to unlock Coxhead's (2000) AWL in first-year coursebooks of NOUN. Coxhead's (2000) academic Word List was used as the baseline word list in this study for its use of computer methods to determine the frequency, range and distribution of words and its utilisation as a resource for EAP researchers and being regarded as more relevant

for EAP learners (I. S. P. Nation, 2001). Using Sinclair's (1991) criteria, the corpus was created with a view to its representativeness, precision and electronic availability.

RESEARCH QUESTIONS

In this study, the academic vocabulary used in the first-year coursebooks for the NOUN is analysed using corpus analysis. The study uses Coxhead's AWL (2000) as a benchmark to examine the following research questions:

1. What is the coverage of frequencies and distributions of AWL items in NOUNC?
2. To what extent are there variations of academic words in NOUNC across the disciplinary fields?
3. What are the most frequent Coxhead's AWL items in the first-year Nigerian ODL coursebooks?

METHOD

CORPUS

A total of 6,802,300 words were collected from three disciplinary fields of e-courseware developed by NOUN. E-courseware was chosen as it is easily accessible, and the electronic format of e-courseware makes the compilation of the corpus easier and faster. In order to convert the digital copies of the coursebooks into a format suitable for analysis, the files were converted to plain text format (.txt) using a computer program called AntFileConverter.

TABLE 2. Word counts in the NOUN coursebook corpus

S/N	Fields	Number of Books	Number of Words
1	Hard Science	75	2,370,502
2	Soft Science	50	2,030,393
3	Non-Science	75	2,401,405
	Total	200	6,802,300

The data were cleaned by removing preliminary pages, references, and tables before conducting the analysis using Anthony's (2021) AntWordProfiler tool. Cleaning the text files was essential for analysing and standardising the data (Benson & Coxhead, 2022). In the next section, the method of data analysis is discussed.

DATA ANALYSIS

To analyse the data, the basic academic vocabulary is selected based on Coxhead's AWL on the basis of frequency principles. The method underlying the frequency principle is the adoption of frequency thresholds "which allowed the maximum number of useful academic words to be recognised" (Benson & Coxhead, 2022, p. 115). This is because the most common criterion for deciding whether a word belongs to a list or not is its frequency of occurrence (Benson & Coxhead, 2022). In compiling the list of the present corpus, Hyland and Tse's (2007) frequency criterion was

used. This is considered a more rigorous and methodological test in which words were categorised as frequent if they occurred above the mean of the overall AWL items in the corpus (Hyland & Tse, 2007).

Thus, the Academic Word List of the National Open University of Nigeria (AWLNOUN) was developed by dividing the corpus into three disciplines and a total of 30 sub-fields. In order to be included in the AWLNOUN, a word must occur at least 20 times in each of the three disciplinary fields and at least 15 times in each of the 30 sub-fields (50% frequency). The current analysis was conducted in accordance with the criteria established by Coxhead (2000), as illustrated in Table 3.

TABLE 3. Summary of word selection criteria of AWLNOUN

Criteria	AWLNOUN
Corpus size	6.8 million-word corpus
Frequency	1132 times on the whole corpus
Range	15 of 30 sub-fields

As shown in Table 3, the corpus size for AWLNOUN is representative of the high frequency and the range. As frequently used in previous studies, the word family was chosen as the unit of counting academic vocabulary in this study (Coxhead, 2000). This is because certain members of a particular word family may be academic while others may not, and vice versa (Benson & Coxhead, 2022). The next section discusses the results of the present study.

RESULTS AND DISCUSSION

This section presents the results with a reference to the relevant and recent literature in a systematic order to answer the following research questions.

THE COVERAGE OF FREQUENCIES AND DISTRIBUTIONS OF AWL ITEMS IN NOUNC

To answer the first research question, a corpus of 6,802,300 was retrieved. Of the total of 6,802,300 running words, 5,336,567 (78.45%) of the corpus were found in the General Service List (GSL). Similarly, 644,407 words were found in Coxhead's (2000) AWL, representing 9.47% of the total tokens, while 821,326 running words that were not included in any list accounted for 12.07% of the corpus. In terms of "various written texts", such as 78.4% in Sutarsyah (2017) and 90.6% in Hirsh (1992), the coverage of the present corpus for GSL words at 78.45% is comparable to the coverage of 78%-92% as predicted by I. S. P. Nation and Waring (1997).

TABLE 4. Coverage of the AWL words in NOUNC

1_GSL_1st_1000.txt	4.940.110	72.62	998
2_GSL_2nd_1000.txt	396.457	5.83	988
3_AWL_570.txt	644.407	9.47	569
Off list	821.326	12.07	
Total	6.802.300	100	

The results presented in Table 4 show that AWL covers 9.47%, while GSL and AWL have a cumulative coverage of 87.92%. According to Coxhead and Nation (2001), a word is considered academic if it occurs frequently in a variety of academic writings and constitutes a substantial part of the vocabulary in an academic corpus. In the academic corpus used for this research, 569 of Coxhead's word families were found in all three fields.

This high coverage shows that academic vocabulary items are important components of the academic texts in the NOUNC corpus. When comparing this study's GSL, AWL, and combined GSL + AWL coverage to other studies, it was found that AWL and the combined coverage of AWL and GSL covered more word lists than found in earlier studies. As a result, variations that highlight the NOUNC corpus's specificity can be seen. Compared to other multidisciplinary corpora, the current corpus's AWL coverage is 9.47%, which is marginally higher than Martinez et al. (2009), E. H. E. Ibrahim et al. (2018b) but marginally lower than that of Hyland and Tse (2007) and Matinparsa et al. (2023).

TABLE 5. Comparing distributions of academic word families with other studies

Word Lists	Coxhead (2000)	Martinez et al. (2009)	Khani and Tazik (2013)	Hyland and Tse (2007)	E. H. E. Ibrahim et al. (2018b)	Matinparsa et al. (2023)	Present Study
GSL	76.1	67.53	76.4	74	58.15	73.78	78.45
AWL	10	9.06	11.96	10.6	8.15	11.46	9.47
GSL + AWL	86.1	76.59	88.00	84.7	66.3	85.24	87.92

In light of the above, the distribution of academic words in the current study is compared with the results of previous studies in Table 5 above. An average of 10% can be seen in most of the results. This confirms the results of previous studies (Coxhead, 2000), in which AWL account for about 10% of an academic text. It should also be mentioned that Hyland and Tse (2007) observed a notable discrepancy in the extent of academic word coverage when they narrowed their attention to specific areas within their corpus. They found that biology had a low coverage rate of 6.2% and computer science had a high coverage rate of 16%. These results demonstrate the validity and high significance of the biology and computer list.

Even though the list in the present study provides a solid overview of the corpus at hand, it is obvious that this coverage is not evenly distributed. For example, since the combination of AWL and GSL does not cover 16% of the words in hard science, students are very likely to be disadvantaged in this area, as they would encounter one unfamiliar word approximately in every four words, rendering the coursebooks incomprehensible. This discrepancy could indicate that hard science writing requires a more specialised and technical vocabulary. This can be attributed to the fact that hard science texts contain more specialised and technical words, as the concepts and phenomena of the hard sciences require precise and specific language or terms.

To further investigate the range, the frequency of each word in the sub-corpora was considered separately. Given the size of this corpus, Coxhead's criteria of frequency — 100 occurrences overall and 10 in each field — seemed surprisingly low. For this reason, a more rigorous and methodological test was applied in which words were categorised as frequent if they occurred above the mean of the total AWL items (i.e., 1132) based on Hyland and Tse's (2007) criteria. Thus, 1132 is used as the minimum threshold to identify which AWL items were used more frequently than average across all disciplines, providing insight into which words might be most broadly useful in academic contexts. Applying this, only 181-word families, i.e. about one-third of the AWL items, could be categorised as frequent according to this criterion (see Appendix

1), with the most frequent words being *process, function, define* and *economic*al. This list is slightly below that of Hyland and Tse's (2007) corpus, in which 192 frequent families were found.

TABLE 6. Frequency, mean and coverage of AWL and GSL

Sub-Corpora	Words (*f*)	AWL Items (*f*)	Mean	AWL Coverage (%)	GSL Coverage (%)	Overall coverage (%)
Hard Science	2370502	218672	384	9.22	75.24	84.46
Soft Science	2030393	217408	382	10.71	79.38	90.09
Non-Science	2401405	208327	366	8.68	80.84	89.52
Overall	6802300	644407	1132	9.47	78.45	87.92

Table 6 shows how the mean value of each field was determined by calculating the total frequency of AWL items in the field and dividing it by the number of AWL items in the corpus (569). Putting these values together resulted in the overall mean value of 1132. Although the highest frequency of words and the highest mean scores are affiliated with hard science, hard science has the lowest overall coverage score due to the compared density in both the AWL and GSL. This resulted in soft science having the highest coverage. Table 7 shows the 60 most frequent academic words against Coxhead's sub-lists.

TABLE 7. 60 frequent academic word families in NOUNC compared with those in Coxhead's AWL sublist families

S/N	Headword	Frequency	Coxhead's Sublist Families	S/N	Headword	Frequency	Coxhead's Sublist Families
1	process	8503	1	31	theory	3927	1
2	function	8439	1	32	period	3912	1
3	define	7692	1	33	compute	3897	2
4	economy	7484	1	34	physical	3831	3
5	individual	6303	1	35	establish	3776	1
6	environment	6083	1	36	legal	3747	1
7	area	5708	1	37	summary	3552	4
8	require	5497	1	38	administrate	3535	2
9	assess	5452	1	39	specific	3531	1
10	communicate	5358	4	40	primary	3528	2
11	involve	5077	1	41	analyse	3393	1
12	create	4887	1	42	issue	3377	1
13	identify	4829	1	43	vary	3334	1
14	concept	4762	1	44	tradition	3273	2
15	structure	4758	1	45	institute	3166	2
16	constitute	4717	1	46	assign	3083	6
17	factor	4614	1	47	achieve	3052	2
18	data	4591	1	48	element	3024	2
19	conclude	4533	2	49	occur	3009	1
20	method	4503	1	50	consume	2969	2
21	source	4391	1	51	cooperate	2901	6
22	major	4371	1	52	affect	2884	2
23	philosophy	4345	3	53	energy	2825	5
24	culture	4198	2	54	available	2776	1
25	principle	4165	1	55	section	2752	1
26	community	4157	2	56	promote	2706	4
27	policy	4076	1	57	authority	2679	1
28	role	4046	1	58	media	2650	7
29	resource	4003	2	59	maintain	2648	2
30	contract	3963	1	60	goal	2643	4

Table 7 depicts that all 60 most frequent word families in NOUNC were found in Coxhead's AWL, albeit in different order of frequency. While thirty-six appeared in Sublist 1, fourteen in Sublist 2, two in Sublist 3, four in Sublist 4, one in Sublist 5, two in Sublist 6, and one in Sublist 7. It is noteworthy to refer to the results in Table 6, which show that all the occurrences in Table 7 are above the overall mean in the corpus. Also found, the thirty-six entries in sublist 1 matched with the entries identified in Hyland and Tse's (2007) multidisciplinary corpus, while ten entries were matched against the list in Martinez et al.'s (2009) agricultural corpus. Driven from the word families shown in Table 7 in accordance with the pertinent literature, it has therefore been suggested that it would be better for university students to concentrate on GSL first. Once these lists have been mastered, AWL should then be the next target (Coxhead & Nation, 2001).

THE VARIATIONS OF ACADEMIC WORDS IN NOUNC ACROSS THE DISCIPLINARY FIELDS

To answer the second research question, the variance of academic words in NOUNC across the fields was calculated. Due to the different objectives and thematic focus of the three fields, the vocabulary used to fulfil the rhetorical and communicative requirements may also differ. The frequency and range in each of the three fields were analysed and found to provide a large proportion of the running words, which are evenly distributed across the text. In each discipline, these word families accounted for an average of 9% of the running words. Table 8 shows that Coxhead's AWLs were present in all three fields, with the proportion being lowest in non-science (8.68%) and soft science being the highest (10.71%).

TABLE 8. Coverage of GSL and AWL across the NOUNC sub-corpora

NOUNC sub-corpora	AWL (%)	GSL first 1000 words (%)	GSL second 1000 words (%)	Total (%)
Hard Science	9.22	68.55	6.69	84.46
Soft Science	10.71	73.96	5.42	90.09
Non-Science	8.68	75.52	5.32	89.52

The coverage of GSL and AWL in the three sub-corpora is shown in Table 8. A closer examination of the distributions within the different sub-fields shows that some items are more common because they are found in one or two particular subfields. Of the 181 families that occurred frequently, only 95 occurred in all three subfields. Word families from the overall most frequently occurring words, such as "*function", "define", "require", "individual", "evaluate"*, and *"include"*, were also found in the top 20 most frequently occurring families in each sub-field, as shown in Table 9.

TABLE 9. Frequently used academic words used in the corpus

Overall				**Hard Science**				**Soft Science**				**Non-Science**			
Family	*f*	Item %	Cum %	Family	*f*	Item %	Cum %	Family	*f*	Item %	Cum %	Family	*f*	Item %	Cum %
process	8503	1.32	1.32	function	4263	1.95	1.95	economy	4702	2.16	2.16	philosophy	3715	1.78	1.78
function	8439	1.31	2.63	process	3750	1.71	3.66	communicate	2904	1.34	3.5	contract	3381	1.62	3.4
define	7692	1.20	3.83	data	3446	1.58	5.24	process	2745	1.26	4.76	constitute	2655	1.27	4.67
economy	7484	1.16	4.99	environment	3118	1.43	6.67	policy	2728	1.25	6.01	define	2564	1.23	5.9
individual	6303	0.98	5.97	area	2838	1.30	7.97	define	2653	1.22	7.23	individual	2381	1.14	7.04
environment	6083	0.94	6.91	define	2475	1.13	9.1	function	2386	1.10	8.33	legal	2197	1.05	8.09
area	5708	0.89	7.80	structure	2351	1.08	10.18	administrate	2376	1.10	9.43	conclude	2031	0.97	9.06
require	5497	0.85	8.65	require	2339	1.07	11.25	individual	2301	1.06	10.49	process	2008	0.96	10.02
assess	5452	0.85	9.50	compute	2337	0.07	12.32	cooperate	2101	0.97	11.46	community	1974	0.95	10.97
communicate	5358	0.83	10.33	energy	2250	1.03	13.35	principle	1906	0.88	12.34	tradition	1906	0.91	11.88

involve	5077	0.79	11.12	factor	2050	0.94	14.29	theory	1840	0.85	13.19	physical	1864	0.89	12.77
create	4887	0.76	11.88	method	2037	0.93	15.22	consume	1830	0.84	14.03	assess	1860	0.89	13.66
identify	4829	0.75	12.63	involve	1945	0.89	16.11	concept	1757	0.81	14.84	create	1848	0.89	14.55
concept	4762	0.74	13.37	assess	1864	0.85	16.96	assess	1728	0.80	15.64	identify	1820	0.87	15.42
structure	4758	0.74	14.11	economy	1781	0.81	17.77	media	1725	0.80	16.44	function	1790	0.86	16.28
constitute	4717	0.73	14.84	vary	1755	0.80	18.57	require	1645	0.76	17.2	primary	1788	0.86	17.14
factor	4614	0.72	15.56	source	1686	0.77	19.34	involve	1644	0.76	17.96	concept	1746	0.84	17.98
data	4591	0.71	16.27	identify	1637	0.75	20.09	major	1640	0.75	18.71	culture	1591	0.76	18.74
conclude	4533	0.70	16.97	individual	1621	0.74	20.83	institute	1622	0.75	19.46	require	1513	0.73	19.47
method	4503	0.70	17.67	occur	1596	0.73	21.56	create	1620	0.75	20.21	involve	1488	0.71	20.18

Taking into account the least frequent words, the distributions remain unequal. It was discovered that 56 families were extremely rare in one subcorpora, 12 in two subcorpora, and 11 in all three subcorpora, using 10% of the mean in each subcorpora as a reference, as recommended by Coxhead (2000). Put differently, 35% of all AWL families have very low occurrence in at least one subcorpora, making them highly unlikely for students to encounter. This result is higher than Hyland and Tse (2007), who found only 27% in their corpus.

THE MOST FREQUENT COXHEAD'S AWL ITEMS IN THE FIRST-YEAR NIGERIAN OPEN AND DISTANCE LEARNING COURSEBOOKS

To answer the third research question, three selection criteria were used: frequency, range and specialised occurrence, adopting Coxhead's criteria. Thus, for a word to be included in AWLNOUN, in addition to being highly frequent, it must be evenly distributed in at least 15 out of the 30 subjects.

TABLE 10. Range of the NOUNC corpus across the subjects

Number of subjects	Number of word families
30	237
29	82
28	84
27	44
26	41
25	13
24	16
23	10
22	10
21	6
20	2
19	2
18	2
17	3
16	2
15	2
Total	556

As shown in Table 10, 565 out of 569 word families met the frequency and range criteria. However, extending the range requirements to only words with a frequency of at least 20 occurrences in each of the three fields consequently led to the exclusion of nine items that did not meet the criterion. This ensures that the final word list consists of items with substantial representation across all three fields. The result was therefore reduced to 556 word families (97.72%) that met the range criteria.

Taking these criteria into consideration, only 181-word families met the threshold, i.e., they were above the overall mean of 1132, had a wide range in the sub-corpora and had specialised occurrence. The results also showed that 56 families appeared in the top 60 in both Coxhead's list and the current analysis, representing the most frequent and consistent words in two the lists. Similarly, the data distributions showed that the majority of words had a wide range across the sub-corpora.

TABLE 11. Range of the AWLNOUN levels in the corpus

Number of subjects	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
30	51	43	16	10	9	3	3	0	1
29	2	3	3	2	1	1	1	1	0
28	0	0	3	2	2	2	1	0	1
27	2	1	0	2	2	2	2	0	0
26	1	3	0	2	2	0	0	0	1
Total	56	50	22	18	16	8	7	1	3

As can be seen in Table 11, these levels are presented based on their frequency of occurrence in the corpus, with level 1 containing the most frequent word families and level 10 containing the least frequent. In this regard, the NOUNC items with the highest frequency levels tend to be well-represented in most subject areas. Levels 1, 2 and 3, for example, had the highest word families, with 56 overall in level 1, followed by 50 in level 2 and 22 in level 3. This proportion extended to level 9. Compared to the first three levels, 1–3, the word families at levels 7, 8 and 9 had the least word families. This turns out to be a more comprehensive result than the corpus of Dang et al. (2017), whose spoken word list was limited to level 4 only.

On the whole, this study has contributed to the field of academic vocabulary research through the analysis of coursebooks used at the NOUN. The findings are useful for EAP practitioners, especially those working in ODL settings. By incorporating such words into the context of course material, content developers in open and distance universities can overcome several challenges related to word distribution identified by Coxhead (2000). This perspective emphasises the importance of the interdisciplinary link between applied linguistics, language teaching, and content creation, and it takes an innovative approach to tailor learning solutions for learners in ODL settings.

Similarly, the study proves worth exploring due to the nature of the student's mode of study. The results of this study showed that familiarity with AWL can increase readers' confidence and improve their understanding of a wide range of texts, including textbooks, scientific journals, and other academic publications. This is consistent with the findings of other researchers (such as Alfraidan, 2010), who found that students' reading comprehension improves when they actively use their AWL to interpret and make sense of what they read. In the same vein, the GSL coverage rate of 78.55% brings the total to 87.92%. Thus, by adding specialised vocabulary, along with words already known, learners can reach the 95% lexical coverage threshold required to understand a text. Given the regularity with which these items appear in NOUN, teachers and students would benefit greatly from spending some time helping students master them.

LIMITATIONS AND IMPLICATIONS

The approach of this study has revealed important implications that should be further elaborated in future research studies. For further studies, several methodological implications can be made. For example, the preparation of the corpus can be achieved through other tools such as Natural Language Toolkit, spaCy, Word2Vec, Stanford CoreNLP, and Corpus Query Processor. Also, this study can be replicated with more content in addition to Coxhead's (2000) AWL. A qualitative research methods design can also be embedded by taking into account the views of coursebook writers. This could present diverse interpretations for further studies.

Regarding the implications, several can be given to the attention of applied linguists. First, when developing university coursebooks, coursebook developers can use AWLNOUN as a reference to facilitate or enhance the presentation of these significant word families due to their high frequency and range. Secondly, teachers no longer have to use their intuition when selecting vocabulary for General Study Unit courses (GSU) but now have quick access to the most commonly used academic words for university coursebook instruction. Besides, the AWLNOUN can help students read academic materials with ease and greater confidence. With 181 word families, the AWLNOUN accounts for 9.47% of the entire corpus. Thus, the AWLNOUN can help students understand enough of the suggested reading materials for ODL to be successful in their chosen fields of study. Thus, the AWLNOUN can be helpful in assisting international students, particularly those majoring in the field of English Language and Applied Linguistics, by giving them a list of terms that are more closely related to their area of study. This kind of word list can prove beneficial in helping these students access the academic language they require to succeed in their courses and any language-related careers they may pursue in the future. Therefore, it can be used for content-based language teaching instruction, language-integrated learning and task-based instruction. For these reasons, both the textbook companies and the language teaching or education departments can show more interest in the field of Applied Linguistics and English as a foreign language, as this study pinpoints the vocabulary development through the corpus.

To conclude, this study prepared and analysed an academic wordlist out of resources used in the Nigerian ODL classes; simply put, it offers a corpus of 6,802,300 words. This corpus is now open for academic purposes to those who want to benefit from it for comparative or contrastive analysis with other academic wordlists. The coverage, frequency, and distributions of the AWL in this corpus provide various insights for comparison with other available corpora. To list a few, the AWLNOUN seems to employ more coverage and distributions of range in comparison with other studies in the pertinent literature (Hyland & Tse, 2007; E. H. E. Ibrahim et al., 2018b; Khani & Tazik, 2013; Matinparsa et al., 2023). Also, the sub-corpora of the NOUNC resulted in the idea that Soft Science offers the richest corpus content in terms of coverage, frequency, and mean scores compared to the AWL and GSL. Finally, this study provides a list of AWL in the NOUN corpus that shows several matches with other corpora and, accordingly, the results of other corpus-based studies (E. H. E. Ibrahim et al., 2018b; Mudraya, 2006; Vongpumivitch et al., 2009). Consequently, this study concludes that the new corpus offers words with richer tokens and words with similar content than other available corpora with specific reference to Coxhead's (2000) AWL.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Malaysia Terengganu and the Ministry of Higher Education, Malaysia, for funding this project (FRGS/1/2020/SSI0/UMT/03/3) under the Fundamental Research Grant Scheme (FRGS) (vote number 59651).

REFERENCES

Alfraidan, A. (2010). *Test taking strategies of EFL Saudi university level learners on two varieties of gap-filling vocabulary achievement tests*. (Doctoral dissertation). University of Essex, the UK.

Anthony, L. (2021). *AntWordProfiler* (Version 1.5.1) [Computer Software]. Waseda University. Available from https://www.laurenceanthony.net/software.

Bahtiar, S. F., Abd Samad, A., Razali, A. B. M., & Ho, Y. W. (2020). A Corpus Analysis of Frequently Occurring Words and their Collocations in High-Impact Research Articles in Education. *3L: Language, Linguistics, Literature*, *26*(4). http://doi.org/10.17576/3L-2020-2604-08

Benson, S., & Coxhead, A. (2022). Technical single and multiword unit vocabulary in spoken Rugby discourse. *English for Specific Purposes*, *66*, 111-130. https://doi.org/10.1016/j.esp.2022.02.001

Brandenburg-Weeks, T., & Abalkheel, A. M. (2021). The Giver: A Corpus-Based Analysis of Word Frequencies. *3L: Language, Linguistics, Literature*, *27*(3). http://doi.org/10.17576/3L-2021-2703-14

Cetinkaya, L., & Sutcu, S. S. (2019). Students' success in English vocabulary acquisition through multimedia annotations sent via whatsapp. *Turkish Online Journal of Distance Education*, *20*(4), 85-98. https://doi.org/10.17718/tojde.640517

Chabbi, Z., & Boukezzoula, M. (2023). *Students' use of academic vocabulary in essay-based exams in the subject of applied linguistics.* (Doctoral dissertation). University of Mohammed Seddik BenYahia, Jijel. Available from http://dspace.univ-jijel.dz:8080/xmlui/handle/123456789/13535

Choo, L. B., Lin, D. T. A., Singh, M. K. M., & Ganapathy, M. (2017). The significance of the academic word list among ESL tertiary students in a Malaysian public university. *3L: Language, Linguistics, Literature*, *23*(4), 56–65. https://doi.org/10.17576/3L-2017-2304-05.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

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

Cunalata Guilcapi, N. M. (2023). *Duolingo Platform and English Vocabulary,* (Master's thesis).

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959-997. https://doi.org/10.1111/lang.12253

Haruna, H. H., Ibrahim, B., Haruna, M., Ibrahim, B., & Yunus, K. (2018). Metadiscourse in students' academic writing: Case study of Umaru Musa Yar'adua University and Al-Qalam University Katsina. *International Journal of English Linguistics*, *8*(7), 83-92. doi:10.5539/ijel.v8n7p83

Haruna, H. H., Tajuddin, A. J. A., & Bashir, I. (2024a). A Corpus-Based Lexical Coverage of Coursebooks in Nigeria: A Case Study. *Register Journal*, *17*, 164-188. Retrieved from https://ejournal.uinsalatiga.ac.id/index.php/register/article/view/836

Haruna, H. H., Tajuddin, A. J. A., & Isa, S. M. (2024b). Vocabulary research trends in applied linguistics through factorial analysis and thematic analysis. *Zhongguo Kuangye Daxue Xuebao*, *29*(3), 91-105. https://zkdx.ch/journal/zkdx/article/view/60

Hirsh, D. (1992). *The vocabulary demands and vocabulary learning opportunities in short novels*. (Unpublished MA Thesis). Victoria University of Wellington, New Zealand.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly*, *41*(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Ibrahim, B., Haruna, H. H., Bashir, I., & Yunus, K. (2018a). The usage of spatial prepositions in the headlines of major Nigerian newspapers. *International Journal of English Linguistics*, *8*(7), 13-22.

Ibrahim, E. H. E., Shah, M. I. A., & Abudukeremu, M. (2018b). A corpus-based lexical study of the frequency, coverage and distribution of academic vocabulary in Islamic academic research articles. *The Journal of Social Sciences Research*, *2*(1),570-577.

Kamasak, R., & Soyaltin, T. E. (2021). Promoting the effectiveness of language teaching through online academic word lists and extensive reading in the covid-19 pandemic. In Montebello, M. (Ed.), *Digital Pedagogies and the Transformation of Language Education* (pp. 165-182). IGI Global Publications.

Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC journal*, *44*(2), 209-232. https://doi.org/10.1177/0033688213488432

Malmstrom, H., Pecorari, D., & Shaw, P. (2018). Words for what? Contrasting university students' receptive and productive academic vocabulary needs. *English for Specific Purposes,* 50, 28-39. https://doi.org/10.1016/j.esp.2017.11.002

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*(3), 183–198. https://doi.org/10.1016/j.esp.2009.04.003

Matinparsa, H., Xodabande, I., Ghafouri, M., & Atai, M. R. (2023). Academic vocabulary in applied linguistics research articles: a corpus-based replication study. *Language Related Research*, *14*(5), 199-221.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235-256. https://doi.org/10.1016/j.esp.2005.05.002

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.

Panmei, B. (2023). Strategic vocabulary learning in vocabulary list learning: insights from EFL learners in Thailand. *3L: Language, Linguistics, Literature*, *29*(1). http://doi.org/10.17576/3L-2023-2901-07

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, *49(*1), 157-165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford University Press.

Sutarsyah, C. (2017). An analysis of student's speaking anxiety and its effect on speaking performance. *IJELTAL, 1*(2), 143-152. http://dx.doi.org/10.21093/ijeltal.v1i2.14

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.

Vongpumivitch, V., Huang, J. Y., & 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

Waluyo, B., & Bakoko, R. (2021). Academic Vocabulary Used by High School Students in Essays and Its Relation to English Proficiency. *3L: Southeast Asian Journal of English Language Studies*, *27*(4). http://doi.org/10.17576/3L-2021-2704-12

Warnby, M. (2022). Academic vocabulary knowledge among adolescents in university preparatory programmes. *Journal of English for Academic Purposes*, *61*(1), 1-24. https://doi.org/10.1016/j.jeap.2022.101203

SUPPLEMENTARY DATA

APPENDIX 1

HEADWORDS OF THE ACADEMIC WORD LIST OF NATIONAL OPEN UNIVERSITY OF NIGERIA (AWLNOUN) FOUND IN COXHEAD'S ACADEMIC WORD LIST

S/N	AWLNOUN	Frequency
1	process	8503
2	function	8439
3	define	7692
4	economy	7484
5	individual	6303
6	environment	6083
7	area	5708
8	require	5497
9	assess	5452
10	communicate	5358
11	involve	5077
12	create	4887
13	identify	4829
14	concept	4762
15	structure	4758
16	constitute	4717
17	factor	4614
18	data	4591
19	conclude	4533
20	method	4503
21	source	4391
22	major	4371
23	philosophy	4345
24	culture	4198
25	principle	4165
26	community	4157
27	policy	4076
28	role	4046
29	resource	4003
30	contract	3963
31	theory	3927
32	period	3912
33	compute	3897
34	physical	3831
35	establish	3776
36	legal	3747
37	summary	3552
38	administrate	3535
39	specific	3531
40	primary	3528
41	analyse	3393
42	issue	3377
43	vary	3334
44	tradition	3273
45	institute	3166
46	assign	3083
47	achieve	3052
48	element	3024
49	occur	3009
50	consume	2969
51	cooperate	2901
52	affect	2884
53	energy	2825
54	available	2776
55	section	2752
56	promote	2706
57	authority	2679
58	media	2650
59	maintain	2648
60	goal	2643
61	feature	2632
62	proceed	2573
63	facilitate	2553
64	select	2405
65	approach	2379
66	research	2362
67	similar	2352
68	access	2307
69	consist	2298
70	normal	2291
71	ensure	2289
72	regulate	2257
73	contribute	2249
74	benefit	2235
75	region	2221
76	distribute	2211
77	respond	2170
78	aspect	2103
79	evaluate	2091
80	component	2073
81	design	2021
82	instance	1997
83	react	1981
84	focus	1964
85	locate	1945
86	participate	1921
87	significant	1916
88	appropriate	1901
89	adequate	1876
90	indicate	1866
91	technique	1852
92	task	1815
93	interact	1771
94	acquire	1763
95	obtain	1726
96	relevant	1707
97	complex	1704
98	conduct	1696
99	objective	1694
100	category	1691
101	assist	1666
102	assume	1659
103	hence	1656
104	final	1655
105	positive	1642
106	consequent	1639
107	potential	1638
108	emphasis	1635
109	range	1632
110	seek	1623
111	equip	1604
112	interpret	1596
113	derive	1590
114	enable	1588
115	aid	1578
116	internal	1574
117	remove	1572
118	capacity	1543
119	negate	1543
120	distinct	1526
121	external	1502

122	link	1501
123	fundamental	1483
124	constant	1478
125	stable	1468
126	impact	1467
127	evident	1465
128	transfer	1460
129	project	1457
130	device	1455
131	concentrate	1455
132	expose	1450
133	construct	1434
134	formula	1429
135	item	1428
136	utilise	1416
137	secure	1416
138	finance	1414
139	document	1408
140	technology	1407
141	professional	1404
142	strategy	1404
143	drama	1382
144	sector	1379
145	transport	1378
146	credit	1365
147	globe	1363
148	convene	1358
149	code	1356
150	logic	1338
151	federal	1338
152	volume	1337
153	job	1334
154	military	1331
155	input	1326
156	implement	1317
157	sex	1311
158	fund	1304
159	adult	1290
160	network	1288
161	challenge	1277
162	psychology	1275
163	legislate	1268
164	status	1265
165	conflict	1248
166	chemical	1245
167	migrate	1227
168	labour	1226
169	commit	1220
170	commission	1211
171	motive	1209
172	file	1206
173	income	1205
174	civil	1203
175	invest	1165
176	image	1161
177	output	1155
178	text	1153
179	compound	1143
180	ethic	1143
181	purchase	1136