Pertanika J. Soc. Sci. & Hum. 33 (6): 2243 - 2262 (2025)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

A Corpus-based Lexical Study of Agricultural Sciences in Open and Distance Learning Coursebooks

Hamisu Hamisu Haruna[1], Azza Jauhar Ahmad Tajuddin[2]*, and Wid Zuriana Sulaiman[3]

[1]*Department of English, Al-Qalam University Katsina, 820102 Katsina, Nigeria*
[2]*Center of Fundamental and Continuing Education, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*
[3]*Marine Conservation and Research Organisation Malaysia (PULIHARA), Lang Tengah Island, 21090 Terengganu, Malaysia*

ABSTRACT

Lexical knowledge is extremely important for learning success, especially in Open and Distance Learning (ODL), where students use printed texts as their main source of learning information. Although studies have indicated a favourable correlation between vocabulary and academic achievement, especially in advanced subjects such as agricultural sciences, recent research has not yet adequately addressed the lexical challenges in agricultural ODL coursebooks. This study involved a corpus analysis of 1.4 million words from five agricultural sub-disciplines at the National Open University of Nigeria (NOUN). Nation's British National Corpus and Corpus of Contemporary American English (BNC/COCA) word lists and Coxhead's Academic Word List (AWL) were used to analyse the lexical coverage required for comprehension thresholds of 95% and 98%. The results show that ODL students require 8,000–9,000 word families for independent reading and that the supplementary lists comprise 1.91% of the required word families. The AWL coverage remained stable across all levels (9.42%-11.75%), indicating a homogeneous academic lexical profile. Limitations include the focus on a single subject area and the reliance on corpus analysis. The findings from this study allow for targeted vocabulary instruction that integrates subject-specific terms, a balanced approach to academic vocabulary acquisition, evidence-based curriculum design, and learner support initiatives. This study makes a valuable contribution by identifying the specific lexical needs of students, creating relevant and comprehensible course materials, and thereby improving the quality of agricultural education in ODL learning environments.

ARTICLE INFO

Article history:
Received: 02 September 2024
Accepted: 17 July 2025
Published: 02 December 2025

DOI: https://doi.org/10.47836/pjssh.33.6.04

E-mail addresses:
hhhamisu@gmail.com (Hamisu Hamisu Haruna)
azzajauhar@umt.edu.my (Azza Jauhar Ahmad Tajuddin)
wid@pulihara.org (Wid Zuriana Sulaiman)
* Corresponding author

Keywords: AntWordProfiler, corpus analysis, coursebooks, English as a Second Language, lexical load, specialised corpus

e-ISSN: 2231-8534

INTRODUCTION

Lexical knowledge is a component that is widely recognised in the academic community as having an impact on text comprehension and overall academic performance, especially in a higher education context (Haruna et al., 2024a; Mugadza et al., 2024; Nation, 2001). However, this relationship becomes even stronger in subject areas where writers tend to use specialised words, such as agricultural sciences, where terms and expressions about the subject area are overused (Hyland & Tse, 2007). The challenge described above is exacerbated in the context of Open and Distance Learning (ODL), where the main form of communication is mainly in written form, making it very important to understand words in their lexical sense (Maphosa & Maphosa, 2023; Yang, 2023).

Although agricultural sciences encompass a wide range of disciplines, sub-disciplines such as fisheries, crop science, etc. entail a variety of lexical variations. Each sub-discipline within the field of study, be it agricultural economics, aquaculture, animal science, agricultural extension, or crop science, has its own reserved terms that go along with the uniqueness of the sub-discipline and encompass concepts, processes, and technologies. These lexical variations pose a major learning problem for students, especially those studying in ODL, as they lack scaffolding. Thus, knowledge of lexical features is crucial, especially in the ODL context where a lack of vocabulary knowledge can hinder comprehension, hinder students' interaction with the course content, and thus affect students' academic success (Haruna et al., 2024b; Townsend et al., 2012).

The agricultural sciences programmes of NOUN, a leading ODL institution, cater to the various needs in the subject area. Normally, it is coursebooks that serve as the vehicle for the delivery of primary knowledge in the virtual classrooms, hence the need to optimise the lexical option. However, while there has been a proliferation of corpus-based lexical studies in various disciplines (Hyland & Tse, 2007), it is evident that there is a dearth of similar studies on ODL texts in agricultural sciences (Bozkurt, 2019). This is worrying, especially when ODL students, who are mostly working and have family responsibilities, rely on self-study using these texts.

Therefore, corpus linguistics, as research of this kind, provides solid and systematic tools for analysing the lexical features of academic text (McEnery & Hardie, 2011). Using word lists based on large corpora such as the BNC/COCA 25000 by Nations (2012), one can determine the vocabulary profile of a particular text to understand the text. Nation (2006) postulated that knowledge of 98% of words is sufficient for unassisted comprehension.

In addition, an important but relatively understudied area of ODL has been identified, namely lexical development, i.e., the increase in lexical demands with an increasing year of study and course level. According to Nagy and Townsend (2012), a good syllabus should show the lexical density of expected academic language

as students progress through the course level. In terms of progression from one course level to the next, such progression is expected in NOUN's four-level agricultural sciences programme (level 200 to 500), but researchers have not gathered empirical evidence for this.

Prompted by these concerns, this research aims to fill these gaps by conducting a corpus-based lexical analysis of Agricultural Sciences Coursebooks (ASCs) in NOUN. The analysis sources were cumulative and comprised large amounts of data: 1.4 million running words from 64 coursebooks in five major sub-disciplines: Agricultural Economics and Agro-business, Aquaculture and Fisheries Management, Animal Science, Agricultural Extension and Management, and Crop Production.

Therefore, this study contributes to the current literature and represents several important advances for the field. First, it provides guidelines for analysing difficulties learners may have in reading and for understanding the lexical complexity of texts used in ODL agricultural sciences courses. Secondly, it highlights specific words that pave the way for targeted learning in this field, and thirdly, by looking at lexical development, researchers provide suggestions for curriculum developers on how to design the development of language density at different levels. Thus, this study not only fills an important gap in the field of corpus-based research but also improves ODL practices in agricultural sciences. As these disciplines become increasingly important for the world's food supply and the ability to feed it sustainably (FAO, 2018 in Kramar et al., 2020), lexical indexing of ODL materials is anything but an intellectual endeavour (Mojtahedzadeh et al., 2024) as it is a method of preparing the next generation of farmers.

LITERATURE REVIEW

Vocabulary Load

Studies on vocabulary, such as Hu & Nation (2000) and Nation (2006), have looked at the vocabulary load—the percentage of known words one must know—for unaided reading comprehension. It refers to a concept in corpus linguistic research that is central to second language acquisition. It has been used as a measure of academic literacy, particularly in the context of text comprehension (Nation 2006; Haruna et al 2024a). It is linked to Laufer's (1989) lexical frequency hypothesis, which states the threshold of word recognition that the reader must reach to comprehend the text. Thus, the concept of 'lexical load' emanates from 'the lexical frequency analysis', in which the number of 'known words' is used as a measure for assessing the difficulty of a text. In this regard, lexical density has emerged as one of the decisive criteria for evaluating texts, and many studies provide evidence for the text difficulty hypothesis. It is therefore important to understand lexical load to determine the vocabulary density required to comprehend written content in academic writing.

In light of the above, readers need to know 98% of the running words to understand a written text without aids,

with one unknown word out of fifty being readable (Nation, 2006). This figure is equivalent to being able to recognise around 8000 to 9000 word families for most written texts. However, for materials such as novels or newspapers, it has been found that the amount practised can go as high as 98-99%, which would require a vocabulary of over 8000-9000 word families (Ibrahim et al., 2018; Kramer, 2023). These results attest to the high level of vocabulary required of readers, particularly in academic settings.

Thus, to calculate the measure of lexical density, it is necessary to use word lists drawn from large, representative samples. One of the earlier attempts to classify vocabulary was made in the West's General Service List (GSL) (1953), but the list is now rather outdated and also small in size, containing only about 2000 word families. The BNC/COCA lists by Nation (2012) contain a vocabulary frequency-based word list with 25000 word families classified according to forms, levels, and frequencies. They are now considered an important tool for carrying out lexical load analyses and provide a more detailed insight into the distribution of words in frequency segments.

Previous Studies on Vocabulary Load

Extensive research has investigated lexical coverage thresholds in different contexts, revealing important insights into the vocabulary requirements for effective reading comprehension. Hirsh and Nation (1992) investigated lexical load using a corpus-based approach on a representative sample of novels and found that the GSL for novels indicates that the reader must recognise about 5000 word families to read them. Although this study provided important initial findings, the analysis was conducted on a rather small corpus compared to the present corpus. Hsu (2014) also used BNC/COCA lists to analyse the complexity of English-language textbooks for engineers. His study found that with 5,000 words, only a comprehension level of 95.5% could be achieved, which was not sufficient for independent reading. Therefore, Hsu called for tutors to use specialised word lists to conduct effective writing tutoring sessions and help engineering students find the vocabulary they need.

Recent studies (such as Haruna et al., 2024c) have used the Nation's BNC/COCA frequency lists to examine the lexical load of learning materials and uncovered information that may be useful in curriculum development and vocabulary pedagogy. In his study, Nguyen (2020) identified the instructionally relevant words in thirty units of English texts used by high school students in Vietnam. The results showed that for better reading comprehension, 95% and 98% of the texts must be understood to attain 3000 and 5000 word families, respectively. Interestingly, these requirements exceeded the students' estimated vocabulary growth to the usual 2000 word families. Additionally, the courses limit the extent to which learners encounter new words in the textbooks, which naturally restricts the learning of these terms.

In addition to Nguyen (2020) and Sun and Dang (2020) investigated the lexical density of a high school series textbook in China. Using Nation's BNC/COCA lists as the basis for the UVLT Webb et al. (2017), the researchers compared the scope of the textbooks with the vocabulary repertoire of 265 high school students. It was found that a word family of 3000 and 9000 existed for 95% and 98% coverage, respectively. Overall, these studies emphasise the role of vocabulary in materials used for language learning and its contribution to students' knowledge. They can also be used as a source of empirical evidence for curriculum development and vocabulary teaching. Considering the mentioned values of key terms and comparing them with lexical features and possible gaps, important recommendations for teachers can be made: It is therefore crucial to develop interventions that promote the use of the most frequent and important words to help children understand academic texts with the help of vocabulary expansion.

Although these studies have contributed much to the knowledge of lexical density in different text types and genres, they also have limitations. Most of them have emerged from earlier word lists, such as the GSL, or have been produced in a very specific domain that does not lend itself well to generalisation. There is also a bias towards exploring the conventional classroom or work-related environment, while new forms such as ODL are excluded or marginalised. The practice of ODL poses particular lexical difficulties for the teaching and learning process. In contrast to face-to-face teaching, where a teacher usually explains only a few unfamiliar terms, ODL students rely mainly on written work (Musingafi et al., 2015; Simpson, 2018). This revealed that there is not much research on lexical load that focuses specifically on the ODL context.

The gap is most obvious when it comes to specialised fields such as agricultural sciences. Since agricultural sciences encompass several subfields, ranging from economics to crop science, the language used in the discipline is both profound and complex (Palanichamy, 2021), due to its specific terminology that can contribute to lexical load. As agriculture is an important sector in the fight against global problems such as hunger (FAO, 2018, in Kramar et al., 2020), ODL materials should be lexically accessible. Nevertheless, the corpus-based study of lexical coverage in ODL coursebooks for agricultural sciences has not yet been conducted.

This study is intended to lexically analyse the Agricultural Sciences Corpus (ASCs) from a major ODL institution in West Africa. Using Nation's (2012) BNC/COCA lists for this large, multidisciplinary corpus, the study would provide, for the first time, empirical evidence of the vocabulary needed in this important field. These findings can be used to assist ODL curriculum designers in designing vocabulary to provide future agricultural professionals with comprehensive access to the knowledge base in their discipline. For clarity purposes, the study investigates

the lexical load of ASCs with 95% and 98% coverage to answer the following research questions:

RQ1: What is the level of difficulty of the NOUN coursebooks for agricultural sciences in terms of vocabulary load, and does this change across levels?

RQ2: How has Coxhead's AWL been integrated into the coursebooks?

RQ3: What are the differences in the coverage of Coxhead's AWL across the levels?

MATERIALS AND METHODS

Research Design

In this study, a quantitative corpus-based research design was used to systematically investigate the lexical load and distribution of academic vocabulary in agricultural sciences coursebooks used in open and distance learning (ODL). The study followed a descriptive-analytical approach and utilised corpus linguistic methodology as the primary research tool. The corpus was designed following the principles of balance, representativeness, and size in corpus linguistics (McEnery & Hardie, 2011). For this analysis, the coursebooks had to meet two selection criteria: first, they had to be core materials, meaning they had to come from the agricultural sciences curriculum at NOUN. Second, the coursebooks had to cover the five major strands of the programme and contain a sufficient variety of texts within these strands to analyse the lexical load in this complex and diverse field. Therefore, texts from such diverse but interrelated areas were included in the corpus to cover the wide variety of terms in the vocabulary of agricultural sciences in areas such as biology, management, and technology.

The study consisted of three phases, which were oriented towards the research questions. Phase 1 profiled the vocabulary in the ASCs using Nation's (2012) BNC/COCA word lists to examine the level of difficulty of the NOUN agricultural sciences course books in terms of vocabulary coverage. Phase 2 examined Coxhead's AWL in ASCs. Phase 3 compared the coverage of ASCs at different levels. The analytical framework was developed to provide evidence-based answers to each research question through systematic quantitative comparisons and descriptive statistical analyses.

The Corpus

The researchers compiled 64 ASCs from five sub-disciplines, totalling around 1.4 million words (Table 1). All coursebooks were downloaded from NOUN e-courseware. Using Anthony's AntFileConverter (2022), all texts that were available in PDF form were downloaded into and converted to txt format. After the file conversion, the data was further cleaned manually by removing preliminary pages, tables, references, captions, and appendices from the documents. The pre-processing of the texts was of crucial importance for the present study, as it helps to normalise the corpus and prepare it for analysis. This is in line with previous empirical studies such as Benson and Coxhead (2022) and Lu and Coxhead (2020).

Table 1
Word counts in the ASC corpus

Sub-disciplines	Number of books	Number of words
Agricultural Economics and Agro-Business	12	301525
Aquaculture and Fisheries Management	14	335208
Animal Science	13	305925
Agricultural Extension & Management	13	280652
Crop Science	12	251542
Total	**64**	**1,474,852**

The difference in the number of books in each subject (12-14 books per subject) is due to the availability and format of these coursebooks under the current system of the National Open University of Nigeria. This slight imbalance is perfectly acceptable and is part and parcel of real-life teaching, which serves to enhance the ecological validity of this study. Also, the inclusion of texts from these different but related fields ensures that the corpus captures the richness of the agricultural lexicon with terms from economics, biology, management, and technology. In compiling the corpus, attention was therefore paid to its representativeness, specificity, full-text availability, and availability in electronic form.

Data Analysis

In this study, the BNC/COCA lists of 25000 words and the other supplementary lists recommended by Nation (2017) were used. AntWordProfiler (Anthony, 2021) is used for corpus-based lexical analyses. When processing the texts, the analysis software generated dense output files describing the tokens and the extent to which the words they spanned corresponded to each frequency level. The above results were used to develop the overall coverage of tokens, an important concept in addressing the vocabulary load issues in the study. These lists were modified and adapted to comprehensively cover all vocabulary in the coursebooks.

The researchers also compared the revised lexical density scores within the different academic levels of the programme. This research aimed to ascertain whether learners' use of challenging words varies as they advance through the educational system. Lastly, the researchers used Coxhead's (2000) AWL to examine the distributions of words in the corpus.

RESULTS AND DISCUSSION

RQ1: What is the Level of Difficulty of the NOUN Coursebooks for Agricultural Sciences in Terms of Vocabulary Load, and does this Change with the Level?

Analysing the vocabulary in the ASCs using Nation's (2012) BNC/COCA word lists provides valuable insights into the lexical demands faced by ODL students in this area. Table 2 shows a nuanced breakdown of coverage by frequency band, revealing both encouraging and worrying trends.

Table 2
Lexical loads of ASC using Nation's (2012) BNC/COCA base word list

BNC/COCA base list	Coverage (%)	Cumulative percentage without supplementary lists	Cumulative percentage with supplementary lists	Examples
1st 1000	65.26	65.26	67	*The, of, and, to, in, a, are…*
2nd 1000	13.94	79.2	80.94	*Product, develop, unit, process…*
3rd 1000	10.71	89.91	91.65	*Agriculture, extension, rural, method…*
4th 1000	2.22	92.13	93.87	*Acid, nutrition, digest, tutor…*
5th 1000	1.33	93.46	**95.2**	*Nutrient, enzyme, elastic, carbohydrate…*
6th 1000	0.83	94.29	96.03	*Poultry, ecosystem, aquatic, secrete…*
7th 1000	0.62	94.91	96.65	*Amino, glucose, soluble, forage…*
8th 1000	0.42	**95.33**	97.07	*Maize, biodiversity, swine, ovary…*
9th 1000	0.25	95.58	97.32	*Lipid, phosphorus, uterus, semi…*
10th 1000	0.24	95.82	97.56	*Semen, fowl, doe, inter…*
11th 1000	0.23	95.05	97.79	*Husbandry, inseminate, lactate, yam…*
12th 1000	0.17	96.22	97.96	*Chromatography, hutch, ovum, faeces…*
13th 1000	0.16	96.38	**98.12**	*Cassava, anova, testes, agribusiness…*
14th 1000	0.15	96.53	98.27	*Ruminant, silage, aquaculture, fructose…*
15th 1000	0.1	96.63	98.37	*Glycogen, hydrolysis, lysine, hypothalamus…*
16th 1000	0.08	96.71	98.45	*Dehydrogenase, methionine, polysaccharide, mammary…*
17th 1000	0.09	96.8	98.54	*Estrous, groundnut, glycerol, spermatozoon…*
18th 1000	0.06	96.86	98.6	*Pullet, rumen, oviduct, follicular…*
19th 1000	0.05	96.91	98.65	*Farrow, analyte, cowpea, evapotranspiration…*
20th 1000	0.05	96.96	98.7	*Allele, oestrus, pyruvate, parturition…*
21st 1000	0.04	97.00	98.74	*Socio, ketone, galactose, gravimetric…*
22nd 1000	0.02	97.02	98.76	*Poult, dehorn, leptospirosis, punnet…*
23rd 1000	0.02	97.04	98.78	*Luteum, cryopreservation, chroma, hexose…*
24th 1000	0.02	97.06	98.8	*Agro, cobby, cuboidal, amniotic…*
25th 1000	0.02	97.08	98.82	*Monosaccharide, maltose, hydrolyte, hygrometer…*
Proper nouns	0.66			*Nigeria, Africa, America, Europe…*
Marginal words	0.49			*E, g, b, c…*
Transparent compounds	0.36			*Livestock, rainfall, runoff, wildlife…*
Abbreviations	0.23			*ATP, LB, WIA, AI…*

The results of the present study in relation to the lexical load of agricultural sciences course books (ASC) based on Nation's BNC/COCA word lists (2012) provided valuable information about the lexical load typical of ODL students. In this study, the first 1,000 most frequently used words covered 65.26% of the given corpus, supporting Nation's (2006) claim about the role of high-frequency words in reading. The second 1000-word families accounted for a further 24.65%, emphasising the role of mid-frequency vocabulary in academic texts. Also, the third 1000-word family had a high percentage of 10.71%. This high percentage may be connected to the high number of agriculture-related terms.

In addition, the study found that the first 5000 most frequently used word families covered only 93.46%, below the expected 95% for minimal comprehension as reported by Laufer and Ravenhorst-Kalovski (2010). This opinion echoes the conclusions of Hsu (2014) on the need to have 5000 words to analyse the frequency of words used in engineering textbooks. This means that it may be difficult for ODL students to understand courses such as agricultural sciences if their vocabulary is limited to 93.46%. However, when the supplementary lists are included, hits reach 95.20% at the fifth 1000-word family level. This corresponds to the minimum threshold formulated by Laufer and Ravenhorst-Kalovski (2010), according to which knowledge of around 5000 word families, together with supplementary lists, should be sufficient to understand the given coursebooks. However, the subsequent frequency ranges do not show a similar proportional increase in coverage, with the 6th to 25th 1000-word families showing a coverage rate of 3.81%. This suggests that while there may be a feature of low-frequency words, their overall contribution to comprehension is not proportionate.

Furthermore, as the study suggests, supplementary lists play a key role in attaining comprehension thresholds, emphasising the role of disciplinary as well as specialised vocabulary. This is consistent with the study by Vuković-Stamatović and Čarapić (2024). This means that the supplementary lists are indispensable for increasing lexical differentiation and the relevance of the corpus of ASCs for teaching. When comparing the cumulative coverage listed in Table 2, one can observe a considerable variation depending on whether the sources of the supplementary lists were included or not. Without the use of supplementary lists, the coverage stops at 96.91% for 19000 word families, which is below the 98% level recommended by Laufer and Ravenhorst-Kalovski (2010) for unaided comprehension. However, when the supplementary lists are added, the overall percentage rises to 98.82%. Thus, analysing the results showed that the effectiveness of these supplementary lists increased by 1.91%. These supplementary lists help to fill the lexical gap by presenting some of the domain-specific terms, technical abbreviations, and multi-word units that are missing from the coursebooks, thus providing learners with the necessary

vocabulary to understand the coursebooks independently.

Comparing the results of this study with the findings presented in the literature, important points can be made. First, the findings support the idea that high-frequency words are generally relevant in different fields, including engineering (Hsu, 2014) and agricultural sciences. This implies that, regardless of discipline, a strong foundation in high-frequency words is crucial for academic success. Second, the study adds to previous research by providing new estimates for the number of word families for efficient reading, requiring 13,000 word families. This shows that agricultural sciences may have more lexical features than other domains, making it necessary for the curriculum designers and teachers in ODL agricultural sciences programmes to invest more time and effort in vocabulary development in this domain. This helps in ensuring fair representation of general academic terms and specialised terms. Such deployment enhances differentiated support to ensure that ODL students understand the subject matter and excel academically in the content area in question. Table 3 shows lexical loads of ASC across the four levels.

Regarding Table 3, which shows the distribution of the lexical load across different levels, 200, 300, 400, and 500, we can make certain assumptions and see the fluctuations and trends. The lexical loads common to each of the four academic levels of 200, 300, 400, and 500 showed varying levels of word complexity. First, the cumulative percentage of coverage in the 200-level text data reached 95% at the 5th 1,000-word family, which is typical of the data for the entire corpus.

At the 200 level, a certain cumulative coverage gets to 95% at the 5th 1000-word family, which again correlates with the overall corpus data. However, in the 300 and 500 levels, this target is achieved at the fifth 1000-word family, while for the 400 level, it is achieved at the sixth 1000-word family. This variation implies that it is possible to find slight differences in the extent to which lexical load differs in the coursebooks at various levels where language is being taught and practiced. This is in line with Sun and Dang's (2020) study on high-school EFL textbooks, which reveals that text simplification can result from form-focussed instructions. However, to attain the 98% optimum threshold, a more distinct pattern is observed. To reach this level in all four levels, students need to know between 8000 and 9000 word families, which are evidently in concordance with the overall assessment of the corpus.

It is essential to note that level 200 has the highest threshold, as 98.52% was only attained at 17000. This is followed by level 400 with 98.01% at 13000, while 300 and 500 levels are slightly behind with 98.14% and 98.02% at 13000 and 10000, respectively. Thus, although there are slight differences in terms of overall words used, the density is slightly higher, which may suggest that students at the highest academic level could benefit from further lexical support. In this respect, these findings are consistent with Nagy and

Table 3
Lexical loads of ASC across the four levels

	Level 200		**Level 300**		**Level 400**		**Level 500**	
BNC/COCA base word	%	Cum %	%	Cum %	%	Cum %	%	Cum %
Proper nouns	0.52	0.52	0.74	0.74	0.54	0.54	0.8	0.8
Marginal words	0.41	0.93	0.56	1.3	0.43	0.97	0.53	1.33
Transparent compounds	0.33	1.26	0.4	1.7	0.32	1.29	0.39	1.72
Abbreviations	0.21	1.47	0.32	2.02	0.21	1.5	0.17	1.89
1st 1000	66.48	67.95	64.48	66.5	65.92	67.42	64.32	66.21
2nd 1000	13.87	81.82	14	80.5	13.45	80.87	14.32	80.53
3rd 1000	9.63	91.45	11.31	91.81	9.68	90.55	12.02	92.55
4th 1000	2.25	93.7	2.17	93.98	2.37	92.92	2.11	94.66
5th 1000	1.3	**95**	1.34	**95.32**	1.39	94.31	1.29	**95.95**
6th 1000	0.83	95.83	0.87	96.19	0.86	**95.17**	0.76	96.71
7th 1000	0.67	96.5	0.56	96.75	0.76	95.93	0.53	97.24
8th 1000	0.42	96.92	0.38	97.13	0.48	96.41	0.41	97.65
9th 1000	0.3	97.22	0.22	97.35	0.32	96.73	0.19	97.84
10th 1000	0.25	97.47	0.25	97.6	0.28	97.01	0.18	**98.02**
11th 1000	0.26	97.73	0.2	97.8	0.29	97.3	0.19	98.21
12th 1000	0.16	97.89	0.18	97.98	0.19	97.49	0.14	98.35
13th 1000	0.15	**98.04**	0.16	**98.14**	0.18	97.67	0.16	98.51
14th 1000	0.17	98.21	0.13	98.27	0.21	97.88	0.1	98.61
15th 1000	0.11	98.32	0.08	98.35	0.13	**98.01**	0.07	98.68
16th 1000	0.09	98.41	0.08	98.43	0.1	98.11	0.05	98.73
17th 1000	0.11	98.52	0.07	98.50	0.13	98.24	0.07	98.8
18th 1000	0.08	98.6	0.04	98.54	0.09	98.33	0.05	98.85
19th 1000	0.05	98.65	0.07	98.61	0.06	98.39	0.04	98.89
20th 1000	0.07	98.72	0.06	98.67	0.08	98.47	0.02	98.91
21st 1000	0.04	98.76	0.04	98.71	0.04	98.51	0.03	98.94
22nd 1000	0.02	98.78	0.02	98.73	0.02	98.53	0.02	98.96
23rd 1000	0.02	98.80	0.02	98.75	0.03	98.55	0.02	98.98
24th 1000	0.02	98.82	0.01	98.76	0.02	98.57	0.02	99
25th 1000	0.03	98.85	0.01	98.77	0.03	98.6	0.01	99.01
Off lists	1.17		1.2		1.38		0.98	

Townsend's (2012) view that appropriate curricula should feature ascending levels of vocabulary complexity. Nonetheless, the differences between levels in the ASC corpus suggest a smoother linguistic differentiation, demonstrating that there is a more gradual progression across the levels.

These findings for lexical load across academic levels have several implications for this study and the previous studies,

both quantitatively and qualitatively. First, the relatively gradual rise in the rate of lexical development in agricultural sciences compared to such disciplines as medicine (Wang et al., 2008) indicates that it may be necessary to use a different strategy in teaching the vocabulary that belongs to the investigated interdisciplinary field. Thus, breaking the learning of new words across the academic spectrum may prove helpful, but it also translates to the fact that advanced learners may require extra help to acquire more specialised vocabulary in a given area. Consequently, the differences between levels suggest that there might be the potential to provide for more purposeful lexical development as learners advance through learning.

Furthermore, Musingafi et al.'s (2015) results regarding ODL students' vocabulary deficiencies highlight the importance of finding ways of remediating weak vocabulary knowledge among distance learners. This is especially important in contexts such as agricultural sciences, for instance, where the percentage of high frequency increases with level, but lexical complexity remains high. From these results, ODL programmes in agricultural sciences should consider implementing vocabulary interventions such as developing specialised words as well as using technology and materials to supplement instructors' real-life word explanations.

RQ2: How Has Coxhead's AWL Been Integrated into the Coursebooks?

The results of Coxhead's (2000) AWL in ASCs provide relevant information about the academic vocabulary in this field. Table 4 summarises the overall analysis of the prevalence of Coxhead's AWL in the corpus.

As shown in Table 4 above, it is crucial to note that GSL comprises a much higher number of general English words, amounting to 76.89% of the corpus. After GSL, AWL has taken a significantly large proportion of 10.53%. Overall, Coxhead's AWL has coverage of 568 word families out of 570. This study is consistent with Coxhead's (2000) observation that the AWL includes many of the most essential words in a wide range of content areas. This shows that the ASC corpus (although slightly higher) aligns with Haruna et al. (2024b) on first-year university coursebooks, Wang et al. (2008) on medical textbooks, and Martínez et al. (2009) in the agrocorpus, which were 9.1% and 9.06% respectively. This indicates that the level of difficulty in these specific domains is comparable. However, the coverage rate of AWL in the

Table 4
Coverage of GSL and AWL in ASC

Word Lists	Token	Coverage %	Word Family	Examples
GSL	1134005	76.89	1895	*The, of, and, to, in, is, a, are, or, for…*
AWL	155236	10.53	568	*Process, method, area, vary, Require…*
Off Lists	185611	12.58	20601	*Rural, objectives, feed, Nigeria…*
Total	**1474852**	**100%**		

ASC corpus is lower compared to 12.3% in Hsu (2014) with respect to engineering textbooks. This may be influenced by the marked disciplinary differences in each field and may be due to the differences in the corpora used in these analyses.

This suggests how relevant AWL in the ODL context is, especially in agricultural sciences, given the marked contribution that AWL makes to the overall vocabulary load. Among the most frequently used AWL words are "process", "method", "area", "economy", and "vary", which indicate the interdisciplinary character of agricultural sciences. It is to be expected that such words occur more than once in each sub-discipline, as they are used to convey concepts and promote academic discussions.

RQ3: What Differences are There in the Coverage of Coursebooks across Different Levels?

To answer this research question, Table 5 compares the coverage of AWL lexical units within the selected text set at different academic levels, such as 200, 300, 400, and 500. It is noticeable that the trends in the coverage of AWL at different levels do not differ drastically, as they range from the lowest 9.42% at the 400 level to the highest 11.75% at the 500 level.

This suggests that the language used in the coursebooks is consistent across the academic levels of the students, indicating that a higher categorisation does not always translate into a larger requirement for the usage of academic vocabulary. Thus, one may predict an increase in the density of academic words at higher levels.

This result contradicts the study by Wang et al. (2008), which found that the coverage of AWL in medical textbooks increases from lower to higher levels. This could explain the observed discrepancy, as agricultural sciences, as a field of study, involves the integration of different areas of academic language, possibly leading to more standardised usage. In this case,

Table 5
Coverage of GSL and AWL in ASC across the four levels

	Level 200	**Level 300**	**Level 400**	**Level 500**	**Total**
GSL tokens	304864	303703	233928	291510	1134005
	77.66%	69.94%	77.11%	76.36%	76.89%
	(26.88%)	(26.78%)	(20.63%)	(25.71%)	100%
AWL tokens	38267	43535	28578	44856	155236
	9.75%	10.96%	9.42%	11.75%	10.53%
	(24.65%)	(28.04%)	(18.41%)	(28.90%)	100%
Off Lists	49449	49915	40849	45398	185611
	12.6%	12.57%	13.47%	11.89%	12.59%
	(26.64%)	(26.89%)	(22.01%)	(24.46%)	100%
Total tokens	392580	397153	303355	381764	1474852
	100%	100%	100%	100%	100%
	(26.62%)	(26.93%)	(20.57%)	(25.88%)	(100%)

it can also be seen that the 500-level has the highest proportion of AWL at 11.75%, suggesting that the frequency of academic words in the final level may increase slightly at best. However, the difference is not very large, and in order to obtain more precise details, a very in-depth analysis is required to determine the exact statistical measure.

It is interesting to note that there is significant engagement with AWL across all levels in the ASC corpus. This could be seen as both an opportunity and a challenge for ODL students of agricultural sciences. On the one hand, the relatively stable nature of academic vocabulary requirements may be an advantage in that it provides opportunities for improved direct vocabulary instruction and intervention, as students are not constantly exposed to significantly different lexical contexts. On the other hand, the lack of enhanced progressions may limit opportunities for the natural development of academic vocabulary through exposure to advanced words in texts (Hyland & Tse, 2007). In this context, and in comparison to conventional face-to-face learning materials, the findings revealed the differences in lexical demand for academic vocabulary across levels.

IMPLICATIONS AND CONCLUSION

This corpus-based analysis shows that lexical accessibility in ODL environments is a critical challenge, while at the same time academic vocabulary demands remain unexpectedly stable across educational levels — findings that call for a fundamental reconceptualisation of vocabulary instruction and course design in distance learning contexts. The study found that the highlighted words of the most frequent 1000 word families provide a fairly comprehensive coverage of 65.26% of the ASC corpus. Nevertheless, this failed to achieve the 95% comprehension rate advocated by Laufer and Ravenhorst-Kalovski (2010). The relevance of supplementary lists was also noted, which increased the overall coverage of the list to 98.82% for optimum reading comprehension.

Moreover, the AWL coverage was found to mean that these words accounted for approximately 10.53%. This result is consistent with the notion that the academic vocabulary of ODL students is overwhelming (Haruna et al., 2024c; Liu, 2023). The study also found that GSL had a larger linguistic scope in the corpus studied (76.89%), but AWL played an important role in capturing subject-specific concepts and processes. This means that lexical density is fairly constant across grade levels, implying that the ASC coursebooks do not show greater compositional density as learners progress. This stable academic vocabulary demand could help to introduce and support ODL students' vocabulary knowledge and usage.

Theoretical Implications

The results of this study make important theoretical contributions to several areas of ODL research, particularly to improving our understanding of lexical accessibility theory and cognitive load theory in self-

directed learning contexts. First, the results provide empirical support for Nation's (2006) vocabulary threshold theory in ODL environments by showing that the 5,000-13,000 word family requirements for independent reading comprehension in agricultural sciences are consistent with theoretical predictions for specialised academic discourse. This finding extends threshold theory beyond the traditional contexts of face-to-face learning and confirms its applicability in distance education, where learners must achieve comprehension autonomously without immediate instructor mediation.

Second, the observed lexical stability across academic levels (AWL coverage 9.42% - 11.75%) challenges the prevailing assumption in academic educational theory that vocabulary complexity should progressively increase with educational progress. This finding contributes to scaffolding theory in ODL by suggesting that consistent lexical demand may actually promote learner autonomy and self-regulation, core principles of distance education theory (Moore, 1993). The stable lexical profile supports Holmberg's theory of guided didactic conversation, in which consistent access to vocabulary enables sustained learner engagement with course materials.

The study advances transactional distance theory (Moore, 1993) by providing quantitative evidence of how lexical factors influence the psychological and communicative space between learners and content in ODL contexts. The high frequency of discipline-specific terms (supplementary lists comprising 1.91% of required word families) shows how specialised vocabulary can either bridge or increase transactional distance, depending on learners' lexical preparation

Finally, the results contribute to the theory of cognitive load (Sweller, 1988) in ODL environments by showing how vocabulary accessibility affects intrinsic cognitive load. The 8,000-9,000-word family requirements suggest that inadequate lexical knowledge may lead to excessive cognitive load that impairs learners' ability to effectively process agricultural concepts in self-directed learning scenarios.

Practical Contributions

The present study has practical contributions for curriculum and pedagogy in ODL contexts. First, more specific approaches to the selection and use of words and the nature of materials and instructions are needed. Since ODL students work through these materials independently without direct teacher support (Haruna et al., 2025; Musingafi et al., 2015), the formal introduction of these terms would contribute to students' overall comprehension and performance.

Specifically, ODL institutions should establish a systematic vocabulary profile when developing courses to ensure that materials provide optimal lexical accessibility while maintaining disciplinary authenticity. Course designers should prioritise the integration of high-frequency academic vocabulary (AWL) at the

beginning of programme sequences, as it is represented at 10.53% at all levels. Similarly, it is crucial to include subject-specific terms in the curriculum design from the very beginning. The strong representation of agricultural terms in all frequency levels shows that curbing vocabulary growth may not have been a one-off thing, but something that should be ongoing and integrated throughout the programme.

Furthermore, ODL institutions should develop tiered vocabulary support systems, including pre-course vocabulary assessments, glossaries aligned to BNC/COCA frequency lists, and multimedia vocabulary learning modules that address the word family threshold identified in this study. Given the stable lexical demands across levels, ODL institutions can implement consistent vocabulary support strategies throughout entire programs rather than increasing complexity progressively. Additionally, the findings suggest that ODL course materials should incorporate explicit vocabulary instruction targeting the gap between GSL (76.89% coverage) and the required comprehension threshold. This could include embedded vocabulary exercises, contextual definition strategies, and spaced repetition systems integrated into digital learning platforms.

Thus, this study contributes to the body of knowledge accumulated by researchers in the field of agricultural and marine sciences and provides a valuable insight into the lexical challenges faced by students in ODL programmes. First, the study provides a guide for curriculum developers and teachers in designing ODL course materials. This, in turn, can lead to a more even distribution of disciplinary knowledge among students, helping to equip the emerging generation of agricultural and marine professionals with the knowledge and tools necessary to address the challenges of today's world.

However, as much as the study has made an important contribution, there are some limitations. The main limitation is that this study was conducted in the field of agricultural sciences, which may not enable extrapolation of the results to other disciplines. In addition, the data for the study were collected through a corpus-based analysis, which is quite reliable but does not take into account the inter-individual variability in students' lexical capacity and learning mechanics. Also, other aspects of language, such as syntactic and discourse patterns, were not analysed in the study for their effects on comprehension. These aspects are important for a proper understanding of lexical demands in the context of academic texts and should be considered in further studies. Similarly, the analysis only looked at lexical load and did not explore other vocabulary knowledge, such as multi-word terms and collocations that could further impede comprehension.

Future research should address these limitations by increasing the size of the corpus to include other ODL institutions and traditional institutions for cross-sectional comparison. In addition, it would be useful to include quantitative data from students and teachers to complement the perceived lexical difficulties and the implementation of current

vocabulary support measures. Furthermore, future research could examine the correlation between the extent of word knowledge, as measured by the number of words, and actual comprehension test scores, as well as the impact of explicitly taught and practised vocabulary on academic performance in ODL agricultural sciences courses. Such studies would not only improve researchers' understanding of lexical demands, but would also serve the best interest of generating research-based, evidence-based information about pedagogy.

ACKNOWLEDGEMENT

This study was conducted as part of the requirements for the Doctor of Philosophy (PhD) degree titled "Investigating Academic Vocabulary in Undergraduate Coursebooks of the National Open University of Nigeria: A Corpus-Based Approach", completed in 2024. The authors gratefully acknowledge the academic support and supervision provided by Universiti Malaysia Terengganu throughout the course of the doctoral research. The authors also wish to express sincere appreciation to the National Open University of Nigeria for granting access to the coursebooks used in this study.

REFERENCES

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

Anthony, L. (2022). *AntFileConverter* (Version 2.0.2) [Computer Software]. Waseda University. Available from https://www.laurenceanthony.net/software.

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.

Bozkurt, A. (2019). From distance education to open and distance learning. In *Advances in educational technologies and instructional design book series* (pp. 252-273). https://doi.org/10.4018/978-1-5225-8431-5.ch016

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

Kramar, I., Marynenko, N., Mischuk, O., Bukhta, V., & Sherstiuk, R. (2020). Economic dimension of digitisation in rural areas. *Engineering for Rural Development 2020 - Proceedings*.

Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. *Ampersand*, *2*, 136-144. https://doi.org/10.1016/j.amper.2015.10.001

Haruna, H. H., Tajuddin, A. J. A., & Isa, S. M. (2024a). Vocabulary research trends in applied linguistics through factorial analysis and thematic analysis. *Zhongguo Kuangye Daxue Xuebao*, *29*(3), 91-105. http://doi.org/10.1654/zkdx.2024.29.3-7

Haruna, H. H., Tajuddin, A. J. A., and Shamsudin, C. M. (2024b). Unlocking academic vocabulary: corpus insights from open and distance English language learning coursebooks. *3L: Southeast Asian Journal of English Language Studies*, *30*(4). *http://doi.org/10.17576/3L-2024-3004-20*

Haruna, H. H., Tajuddin, A. J. A., & Bashir, I. (2024c). A corpus-based lexical coverage of coursebooks in Nigeria: A case study. *Register Journal/Register*, *17*(1), 164-188. https://doi.org/10.18326/register.v17i1.164-188

Haruna, H. H., Iftikhar, S., Isa, S. M., & Hasan, M. M. (2025). Effect of professional development, collaborative learning, and technology integration on teachers' self-access support competence.

Studies in Self-Access Learning Journal, 16(3), 526–565. https://doi.org/10.37237/160304

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696. https://doi.org10.4236/als.2022.102016

Holmberg, B. (2005). *Theory and practice of distance education*. Routledge. https://doi.org/10.4324/9780203973820

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, *33*, 54-65. https://doi.org/10.1016/j.esp.2013.07.001

Hu, M. and Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13(1): 403-430.

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. (2018). The usage of spatial prepositions in the headlines of major Nigerian newspapers. *International Journal of English Linguistics*, *8*(7), 13-22. https://doi.org/10.5539/ijel.v8n7p13

Kramer, B., (2023). *The vocabulary of extensive reading: A corpus analysis of graded readers*. [Doctoral dissertation, Temple University]. ProQuest Dissertations Publishing. https://www.proquest.com/dissertations-theses/vocabulary-extensive-reading-corpus-analysis/docview/2858132623/se-2

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In Lauren and Nordman (Ed.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Multilingual Matters LTD Clevedon.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15-30. http://hdl.handle.net/10125/66648

Liu, C. (2023). A corpus-based study of vocabulary in massive open online courses (MOOCs). *English for Specific Purposes*, 72, 40-50. https://doi.org/10.1016/j.esp.2023.08.002

Lu, C., & Coxhead, A. (2020). Vocabulary in traditional Chinese medicine: Insights from corpora. *ITL-International Journal of Applied Linguistics*, *171*(1), 34-61. https://doi.org/10.1075/itl.18020.lu.

Maphosa, V., & Maphosa, M. (2023). Opportunities and challenges of adopting MOOCs in Africa: A systematic literature review. In *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.1001298

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

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Mojtahedzadeh, R., Mohammadi, A., & Zarei, A. (2024). The impact of open education on health professional training: From open educational resources to MOOCs. In *Advances in Science, Technology & Innovation/Advances in science, technology & innovation* (pp. 163-171). https://doi.org/10.1007/978-3-031-51244-5_18

Moore, M. G. (1991). Distance education theory. *American Journal of Distance Education*, 5 (3), 1-6. https://doi.org/10.1080/08923649109526758

Mugadza, J., Kilag, O. K., Uy, F., Colina, G., Abella, J., & Rabi, J. I. I. (2024). Vocabulary learning: Approaches aligned with common core standards. *International Journal of Qualitative Research for Innovation, Sustainability, and Excellence (IJQRISE)*, *1*(2), 20-27.

Musingafi, M. C., Mapuranga, B., Chiwanza, K., & Zebron, S. (2015). Challenges for open and distance learning (ODL) students: Experiences from students of the Zimbabwe Open University. *Journal of Education and Practice*, *6*(18), 59-66.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91-108. https://doi.org/10.1002/rrq.011

Nation, I. S. P. (2001). *Learning vocabulary in another language* (Vol. 10, pp. 126-132). Cambridge: Cambridge university press.

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2012). What does every ESOL teacher need to know? Closing plenary address at the 2012 CLESOL conference in Palmerston North.

Nation, P. (2017). How vocabulary is learned. *Indonesian JELT*, *12*(1), 1-14. http://dx.doi.org/10.25170/ijelt.v12i1.1458.

Nguyen, C. (2020). Lexical features of reading passages in English-language Textbooks for Vietnamese high-school students: Do they Foster both content and vocabulary gain? *RELC Journal*, *52*(3), 509-522. https://doi.org/10.1177/0033688219895045

Palanichamy, Y. (2021). A *topic modeling assessment of emerging research trends in the environmental science and engineering discipline*. [Unpublished master's thesis]. Ryerson University Canada.

Simpson, O. (2018). *Supporting students in online, open and distance learning*. Routledge.

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, *93*, Article 102279. https://doi.org/10.1016/j.system.2020.102279

Sweller, J. (1988). Cognitive load during problem solving, *Cognitive Science,* 12, 257-285.

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. https://doi.org/10.1086/663301

Vuković-Stamatović, M. and Čarapić D. (2024). Vocabulary profile, lexical density and speech rate of science podcasts: How adequate are they for EAP and ESP listening? *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, (47), 201-226. https://doi.org/10.17398/2340-2784.47.201

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*(4), 442-458. https://doi.org/10.1016/j.esp.2008.05.003

Webb, S., Sasao, Y., & Balance, O. (2017). The updated vocabulary levels test. *ITL-International Journal of Applied Linguistics*, *168*(1), 33-69 https://doi.org/10.1075/itl.168.1.02web.

West, M., (1953). *A general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology.* Harlow.

Yang, L. (2023). An "Interactive Learning Model" to enhance EFL students' lexical knowledge and reading comprehension. *Sustainability*, *15*(8), Article 6471. https://doi.org/10.3390/su15086471

APPENDIX

Supplementary Table
100 most frequent academic word family headwords in ASC

ASC	Freq.	ASC	Freq.	ASC	Freq.	ASC	Freq.
process	3060	period	1089	benefit	767	remove	549
method	2599	evaluate	1060	role	758	communicate	544
area	2450	labour	1057	finance	751	consist	539
economy	2347	principle	1014	ensure	745	job	530
vary	2338	energy	1001	assign	739	invest	527
require	2160	income	989	research	728	create	525
environment	2096	affect	984	concept	726	range	523
resource	2037	structure	977	obtain	720	react	523
function	1916	occur	944	physical	713	credit	521
factor	1903	output	943	culture	703	margin	519
involve	1734	specific	939	utilise	700	participate	515
analyse	1670	major	937	significant	698	cycle	510
individual	1522	select	871	normal	674	implement	510
define	1492	summary	870	adequate	668	estimate	492
community	1468	facilitate	868	establish	668	indicate	491
identify	1417	distribute	834	maximise	654	task	490
consume	1314	technique	828	alternative	631	element	487
source	1305	proceed	827	strategy	621	objective	476
available	1234	approach	819	cooperate	566	assume	475
achieve	1233	institute	813	layer	564	locate	469
project	1174	theory	812	policy	563	similar	447
input	1160	impact	800	component	560	commodity	446
conclude	1113	data	788	administrate	556	primary	445
goal	1112	maintain	787	equip	556	region	442
assess	1093	design	772	chemical	552	promote	439