



UNIVERSITY OF LEEDS

This is a repository copy of *Vocabulary in academic lectures*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/193067/>

Version: Accepted Version

Article:

Dang, TNY orcid.org/0000-0002-3189-7776 (2022) Vocabulary in academic lectures. Journal of English for Academic Purposes, 58. 101123. ISSN 1475-1585

<https://doi.org/10.1016/j.jeap.2022.101123>

© 2022 Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Citation of published version:

Dang, T. N. Y. (2022). Vocabulary in academic lectures. *Journal of English for Academic Purposes*. 58, Special issue: Lecture discourse and lecturer training. <https://doi.org/10.1016/j.jeap.2022.101123>

Pre-publication version

Vocabulary in academic lectures

Thi Ngoc Yen Dang

University of Leeds

ABSTRACT

Vocabulary is essential for comprehending lectures in university courses. No studies have explicitly compared vocabulary in lectures from different contexts. In this study, three corpora were developed from university lectures in EMI courses, non-EMI courses, and open access EMI courses. Analysis of the vocabulary load showed that lectures in non-EMI courses were more lexically demanding than open access EMI courses, which were more lexically demanding than EMI courses. Analysis of the coverage of Dang et al.'s (2017) Academic Spoken Word List (ASWL) showed that compared to general words, learning ASWL words is a shortcut for learners to improve their lecture comprehension. Moreover, while knowledge of items from all ASWL levels could benefit learners from the three contexts, attention to Levels 1-3 items would probably be more beneficial for those studying in EMI courses and open access EMI courses. Attention to Level 4 items appeared to be more useful for those studying in non-EMI courses.

Key words: *lexical coverage; vocabulary; lecture; academic listening, word list; EMI*

1. Introduction

To achieve academic success in university courses, students need to understand lectures. However, many students whose first languages (L1s) are not English find academic listening challenging due to their insufficient knowledge of specialized vocabulary (Evans & Green, 2007; Evans & Morrison, 2011; Mulligan & Kirkpatrick, 2000). Research also shows that among various domain-specific knowledge and domain-general cognitive abilities, vocabulary

knowledge is the strongest predictor of academic listening comprehension (Wallace, 2020) and that listening comprehension increases according to the percentage of known words in the input (Durbahn et al., 2020; van Zeeland & Schmitt, 2013a). Therefore, investigating the vocabulary in lectures would provide useful insights into the lexical challenges faced by English as an additional language (EAL) students when comprehending lectures and the extent to which existing academic word lists could help learners to deal with this challenge. This would in turn help EAL students to improve their academic listening comprehension. In the last two decades, English has been widely used as the medium of instruction in university courses around the world, not just in English-speaking countries (Dearden, 2014). Moreover, apart from delivering lectures to regular on-site students in attendance, some universities in English-speaking countries (e.g., MIT, Yale, and Stanford) also make the recordings of these lectures and course materials freely available online for international audiences from various first languages (L1s) and socio-economic backgrounds to promote open education. Given the diverse contexts in which lectures are delivered, EAP (English for Academic Purposes) teachers, who prepare EAL students for their study in English-medium courses, would probably benefit most from information related to their specific contexts. Yet no studies have explicitly compared the vocabulary in academic lectures in different contexts. To address this gap, this study investigates vocabulary in academic lectures from English-medium (EMI) courses, non-EMI courses, and open access EMI courses. Vocabulary was examined from two perspectives: (a) the vocabulary load of academic lectures and (b) the extent to which Dang et al.'s (2017) Academic Spoken Word List helps learners to deal with this load.

1.1. Specialized vocabulary

According to Nation (Nation, 2013, 2016), specialized vocabulary can be classified into academic vocabulary and technical vocabulary. Academic vocabulary occurs in a wide range of academic disciplines while technical vocabulary occurs frequently in a specific discipline or topic. For example, *align*, *optimize*, and *evaluate* are academic words whereas *malignant*, *tumour*, and *cartilage* are technical words in medicine. Research has shown that knowledge of specialized vocabulary is closely related to academic performance (Coxhead, 2020). That is, the more specialized vocabulary learners know, the better their academic performance is. There are different views about whether technical vocabulary or academic vocabulary should be the focus of EAP programs (Nation, 2016). The first view suggests that EAP programs should focus on

technical vocabulary. In this way, learners will be introduced to the lexical items that are the most relevant to their specific disciplines. As a result, they will see clearly the link between what they have studied in EAP courses and what they need to do in their subsequent academic study. This would then motivate them to learn vocabulary. This approach promotes the development of technical word lists in engineering (Ward, 2009), medicine (Dang, 2020a; Watson-Todd, 2017), hard sciences (Dang, 2018a), and soft sciences (Dang, 2018b).

By contrast, the second view suggests that EAP programs should focus on academic vocabulary because it has wider applications. This approach is relevant to EAP programs in which students come from various disciplines or teachers lack background knowledge of students' specific disciplines (Dang et al., 2017). The focus on academic vocabulary is also supported with evidence from previous studies. Corpus-driven studies have consistently shown that academic vocabulary makes up a substantial percentage of words in academic written and spoken texts (e.g., Coxhead, 2000; Dang et al., 2017; Gardner & Davies, 2014) whereas intervention studies have revealed that academic vocabulary was less likely to be learned incidentally by second language (L2) learners than technical vocabulary (Vidal, 2003, 2011). This approach justifies the development of academic word lists such as the Academic Word List (Coxhead, 2000), the Academic Vocabulary List (Gardner & Davies, 2014), and the Academic Spoken Word List (Dang et al., 2017). Despite these different viewpoints, researchers have suggested that specificity in EAP research should be treated with flexibility and consideration of learners in a particular context (Hyland, 2016), and vocabulary studies should state clearly their target learners (Nation, 2016). As academic vocabulary has wider applications, the present study will focus on this type of specialized vocabulary.

1.2. Vocabulary in academic lectures

Academic vocabulary has received a great deal of attention from both vocabulary and EAP researchers. However, these studies have mainly focused on vocabulary in academic written discourse such as research articles, university textbooks, and student writing (Coxhead, 2000; Gardner & Davies, 2014). Few studies have examined vocabulary in lectures, possibly due to the challenge of collecting and analyzing spoken data (O'Keeffe et al., 2007). However, with the availability of academic spoken corpora (e.g., the British Academic Spoken English and the

Michigan Corpus of Spoken English) and open access lectures, there is an increase in the number of studies investigating vocabulary in lectures.

1.2.1. Vocabulary load of lectures

One trend in research on vocabulary in lectures is vocabulary load. A typical approach to determine vocabulary load is to estimate the number of known words in texts. The larger the percentage of known words, the lighter the vocabulary load. This is likely to enhance comprehension, although lexical knowledge cannot guarantee that a listener will always recognize a lecturer's intended meaning. Based on empirical studies with L2 speakers of English, van Zealand and Schmitt (2013) found that the amount of coverage needed for listening comprehension ranged from 90% to 100% with certain lexical coverage cut-off points leading to certain degrees of comprehension. 90%, 95%, 98% and 100% lexical coverage led to comprehension of 73.5%, 76.5%, 82.2%-100%, and 96.2%-100%, respectively. They then suggested that 90% coverage was likely to result in 'relatively high' comprehension rates; 95% coverage to 'relatively high and stable' comprehension rates; 98% coverage to 'very high' comprehension rates; and 100% to 'almost full' comprehension rates (p.474). Although higher coverage is likely to lead to better comprehension, achieving 100% coverage is too ambitious a goal for L2 learners. Thus, subsequent lexical coverage studies usually adopt 90%, 95%, and 98% to indicate different degrees of comprehension (see Nurmukhamedov & Webb, 2019 for the timeline of these studies).

Only Dang and Webb (2014) has attempted to determine the vocabulary load of lectures and seminars. Their analysis of the vocabulary in the British Academic Spoken English (BASE) corpus showed that 2,000 word-families, 4,000 word-families, and 8,000 word-families were needed to achieve 90%, 95%, and 98% coverage, respectively. A word-family includes a base forms, its inflected forms, and derived forms. For example, *discuss*, *discussed*, *discusses*, *discussing*, *discussion*, *discussions*, and *undiscussed* are considered as one word-family. While their analysis provided useful information about the vocabulary load of lectures, they did not distinguish between lectures and seminars in their analysis. As the linguistic features may vary according to the kind of speech event (Biber, 2006), the vocabulary load of academic lectures remains to be determined. Moreover, the lectures in Dang and Webb's study were from non-EMI courses delivered solely to on-site students. It is unclear whether the same results would be found

in lectures in EMI courses and lectures in non-EMI courses which are publicly available to wider audiences. Addressing this issue would allow EAP teachers to achieve a better understanding of the vocabulary load faced by their students when listening to lectures in different contexts so that they could provide their students with relevant support.

1.2.2. Coverage of academic word lists in lectures

Another trend in research on vocabulary in lectures is developing academic word lists which capture the words that learners will encounter frequently in academic lectures so that they can better deal with the tasks in their subsequent academic studies (Coxhead, 2020). Several academic word lists have been developed for EAP learners. Coxhead's (2000) Academic Word List (AWL) is probably the best known. Developed from an academic written corpus of 3.5 million words, the AWL consists of 570 word-families. This list consistently covers around 10% of the words in academic written texts (see Coxhead, 2011, 2016 for a review), but only 4.6% of the words in lectures (Thompson, 2006), 4.41% of lectures plus seminars (Dang & Webb, 2014), 3.56% of tutorials (Coxhead & Dang, 2019), and 2.52% of labs (Coxhead & Dang, 2019). The lower coverage of the AWL in academic spoken English than in academic written English is not surprising because lexical differences between the two modalities have been documented by Biber (2006) and Dang (2020b).

To better capture vocabulary in academic spoken English, Dang et al. (2017) developed an Academic Spoken Word List (ASWL) of 1,741 word-families from a 13-million word corpus which represents lectures, seminars, labs, and tutorials from 24 academic subject areas. The list covered around 90% of the words in lectures, seminars, labs, and tutorials combined. One innovative feature of the ASWL is its division into levels based on the frequency of the ASWL words in general English. This approach helps it match the needs of EAP learners at different language proficiency levels. Level 1 consists of ASWL word-families that occur frequently in academic spoken English but are also among the most frequent 1,000 word-families of general English (e.g., *know*, *see*). Level 2 represents ASWL words that occur frequently in academic spoken English but also belong to the 1,001st-2,000th most frequent items of general English (e.g., *therefore*, *determine*). Level 3 includes ASWL words that occur frequently in academic spoken English but are also among the 2,001st and 3,000th most frequent items of general English (e.g., *achieve*, *significant*). Level 4 represents the ASWL words that occur frequently in

academic spoken English but are outside the most frequent 3,000 words of general English (e.g., *straightforward, optimize*). Learners can focus on the ASWL levels that are relevant to their current knowledge of general vocabulary and may potentially recognize 92%-96% of the words in lectures, seminars, labs, and tutorials combined. Later, Coxhead and Dang (2019) found that this list covered more than 92% of the words in tutorials and nearly 91% of the words in labs. However, they did not specifically examine the coverage of the ASWL in lectures, a key component in academic study, nor did they consider the possible effect of different lexical delivery contexts on vocabulary load. Given the diverse contexts in which EAP students are likely to study, examining the potential coverage provided by the ASWL in lectures delivered in different contexts would provide EAP teachers with further insight into the value of this list for learners in different contexts.

1.3. Research questions

Given the lack of research explicitly comparing the vocabulary in academic lectures from different contexts, this study will find the answers to the following research questions:

1. What is the vocabulary load of academic lectures delivered in English-medium instruction (EMI) courses, non-EMI courses, and open access EMI courses?
2. To what extent can knowledge of Dang et al.'s (2017) Academic Spoken Word List move learners with different vocabulary levels to different levels of comprehension of lectures in each context?

2. Methodology

2.1. Corpora

Three corpora¹ were developed for this study: an EMI corpus, a non-EMI corpus, and an open access non-EMI corpus. The classification of these corpora is based on the lecture settings and students' L1 status (see Table 1).

Table 1. Description of the lecture corpora

Features	EMI corpus	Non-EMI	Open-access non-EMI
Setting	regular lecture with regular on-site students	regular lectures with regular on-site	Most sessions are regular lectures with regular on-site students in

	in attendance	students in attendance	attendance but recorded to make publicly available for international audience; lecturers are aware of this fact. Some were pre-recorded lectures, whose scripts had been prepared in advance.
Country of delivery	Hong Kong and Finland	New Zealand, UK, and Ireland	US
Students' L1 status	mixed, mainly Chinese (HKCSE) and Finnish (ELFA)	mixed	mixed; wider audiences than those in the other corpora

The EMI corpus was created from the transcripts of the lectures in the Hong Kong Corpus of Spoken English (HKCSE), which were recorded in Hong Kong, and the lectures in the corpus of English as a Lingua Franca in Academic Settings (ELFA), which were recorded in Finland. Although English is one of the official languages in Hong Kong, most of the students in the HKCSE were Chinese L1 speakers; therefore, lectures from this corpus were classified as EMI. The non-EMI corpus was derived from the transcripts of lectures recorded at a university in New Zealand and the lecture components in the British Academic Spoken English Corpus (BASE), the Michigan Corpus of Academic Spoken English (MICASE), the Limerick-Belfast Corpus of Academic Spoken English (LIBEL), and the Pearson International Corpus of Academic English (PICAIE). The open-access non-EMI corpus was developed from the transcripts of lectures from open courses delivered by Yale University, Massachusetts Institute of Technology, Stanford University, and the University of California at Berkeley. The links of the publicly available corpora whose texts were used to develop the corpora in the present study are listed in Appendix 1.

Table 2 shows that each of the three corpora has two subcorpora representing hard subjects (e.g., Astronomy, Biology, Health Science, and Computer Science) and soft subjects (e.g., Linguistics, Art, Law, and Business). The hard-soft classification is based on Becher's (1989) classification of academic subjects in higher education. This classification was chosen because it has been validated in various studies and was adopted in the development of the BASE and the British Academic Written Corpus. Due to the availability of transcripts, the EMI corpus was much

smaller than the other two corpora and only sampled lectures from Hong Kong and Finland. Moreover, the hard and soft subcorpora were not balanced. Given these limitations, the findings of this study are only exploratory.

Table 2. Sizes of the lecture corpora (words)

EMI courses		Non-EMI courses		Open access EMI courses	
Hard	172,944	Hard	4,772,665	Hard	7,880,005
Soft	80,962	Soft	6,672,340	Soft	8,182,275
Total	253,906	Total	11,445,005	Total	16,062,280

2.2. Analysis

Heatley et al.'s (2002) RANGE program was used to analyze the vocabulary in each corpus. To determine the vocabulary load of lectures, Nation's (2012) 25 1,000- BNC/COCA word lists were used as the base word lists. These word lists represent the 1st -25th most frequent 1000 word-families in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The cumulative coverage provided by each 1,000 word-family level of the BNC/COCA lists (from the most frequent to the least frequent) was gradually added up until the 90%, 95%, and 98% coverage points were reached. Coverage provided by proper nouns (e.g., *Alexander, Elvis*), marginal words (e.g., *ahhhhh, hahaha*), compounds (e.g., *bedroom*), abbreviations (e.g., *ASEAN*) was also added to the cumulative coverage because these words are likely to pose a lighter learning burden than other words (Nation & Webb, 2011).

To identify the coverage of the ASWL in the lecture corpus, the ASWL levels, which were graded based on the BNC/COCA lists, were used as the base word lists. The potential coverage of academic lectures that learners with different vocabulary levels may achieve with the support from the ASWL was the sum of the coverage provided by (a) the word-families at learners' current vocabulary levels, (b) word-families at the ASWL levels that are beyond their current vocabulary knowledge, (c) proper nouns and (d) marginal words.

3. Results

RQ1. What is the vocabulary load of academic lectures delivered in different contexts?

The coverage of each 1000 BNC/COCA word level in the three corpora is presented in Appendix 2. Table 3 presents the cumulative coverage of the 1,000-word frequency levels together with proper nouns, marginal words, compounds, and abbreviations. Knowledge of the most frequent 2000 word-families was necessary to reach 90% coverage of academic lectures regardless of the context. The differences in the lexical demands of lectures in the three contexts were greater when the 95% and 98% coverage cut-off points were examined. To reach the 95% coverage, a larger vocabulary size was needed in the case of non-EMI courses (the most frequent 4,000 word-families) than in the case of open access non-EMI courses and EMI courses (the most frequent 3,000 word-families). If the 98% coverage was set, the difference in the vocabulary load of lectures across the three contexts was even larger: 9,000 word-families (non-EMI), 8,000 word-families (open-access non-EMI), and 7,000 word-families (EMI).

Table 3. Number of word-families needed to reach certain lexical coverage of lectures in different contexts

Number of word-families	EMI	Non-EMI	Open-access non-EMI
1,000	83.43	84.53	84.99
2,000	90.61	90.64	91.21
3,000	95.45	94.88	95.45
4,000	96.57	96.23	96.61
5,000	97.40	96.97	97.26
6,000	97.80	97.40	97.67
7,000	98.10	97.69	97.96
8,000	98.29	97.92	98.16
9,000	98.40	98.08	98.29
10,000	98.52	98.22	98.42
11,000	98.59	98.33	98.53
12,000	98.65	98.42	98.60
13,000	98.77	98.48	98.66
14,000	98.83	98.52	98.71
15,000	98.87	98.57	98.75
16,000	98.91	98.61	98.78
17,000	98.94	98.63	98.81
18,000	98.96	98.65	98.83

19,000	98.98	98.67	98.85
20,000	99.02	98.68	98.86
21,000	99.03	98.69	98.87
22,000	99.03	98.70	98.88
23,000	99.03	98.71	98.89
24,000	99.03	98.71	98.89
25,000	99.03	98.72	98.90

RQ2. To what extent can knowledge of the Academic Spoken Word List move learners with different vocabulary levels to different levels of comprehension of lectures in each context?

Table 4 shows that the coverage of the ASWL in the three contexts was fairly similar.

Irrespective of the contexts, the whole list covered from more than 87% to nearly 90% of the words in academic lectures. Level 1 always provided the largest coverage, followed by Levels 2 and 3. Level 4 provided the lowest coverage.

Table 4. Coverage of each level of the Academic Spoken Word List (Dang et al., 2017)

	Number of words	EMI	Non-EMI	Open access non-EMI
Level 1	830	76.83	80.47	81.29
Level 2	456	6.40	5.13	5.12
Level 3	380	3.29	2.64	2.90
Level 4	75	0.51	0.19	0.51
Total	1741	87.03	88.43	89.82

The analysis also revealed there are certain variations in the most frequent lexical words in each context. Appendices 3-6 list the top 50 ASWL lexical words at Levels 1-4 in each lecturing context (EMI, non-EMI, and open access non-EMI). Table 5 presents the number of unique items in each context (e.g., the items only appearing in the top 50 ASWL lexical words in one context, but not in the other two contexts). In terms of Levels 1 and 2, lectures in EMI courses had the largest number of unique items while the number of unique items in lectures in the non-EMI courses and in open access non-EMI courses were similar. Yet in terms of Level 3, open access EMI courses had the smallest number of unique items whereas lectures from the other two contexts had a similar number of unique items. In terms of Level 4, non-EMI courses had the largest number of unique items while a similar number of unique items were found in the other

contexts. The variation in the ranking of the top ASWL items indicates that the relative value of each ASWL word might vary according to context. However, the analysis also showed that all ASWL words occurred in all three corpora. This suggests that while there are differences in the relative value of each item, overall the ASWL is a useful list for learners to deal with lectures in a wide range of contexts.

Table 5. Items unique to the top 50 lexical ASWL items in each context

ASWL levels	EMI	Non-EMI	Open-access non-EMI
Level 1	10	3	4
Level 2	24	17	17
Level 3	23	23	17
Level 4	5	7	5

Table 6 shows that depending on learners' prior vocabulary knowledge, learning an extra 75 to 1,741 word-families from the ASWL would enable them to achieve 91%-96% coverage of academic lectures in different contexts. The first three rows show that the potential coverage provided by the ASWL for learners with insufficient knowledge of the most frequent 3000 word-families in the three lecturing contexts was fairly similar. Learners who have not mastered the most frequent 1,000 word-families would be able to achieve 91%-94% coverage of lectures if they learn only 1,741 word-families from the ASWL. Similarly, learners with the vocabulary level of the most frequent 1,000 word-families would achieve around 92%-94% coverage of lectures if they learn 911 ASWL word-families that are beyond their vocabulary knowledge.

In contrast, as can be seen from the second row of Table 3, learning the most frequent 2,000 word-families of general vocabulary could potentially allow learners to reach only 90% coverage of lectures in different contexts. Compared to learning words from the next frequency level of general English, learning the ASWL word-families that are beyond their current vocabulary level means that learners with insufficient knowledge of the most frequent 2,000 word-families of general English need to study a smaller number of words (911 or 1741), but they can achieve higher lexical coverage (91%-94%). The third row of Table 6 presents the potential coverage gained by learners with the vocabulary level of the most frequent 2,000 word-families if they learn the ASWL. Studying 455 ASWL word-families that are beyond their current vocabulary level would allow them to achieve nearly 95% coverage of academic lectures. This amount of

coverage is around the same coverage achieved from learning 1,000 word-families from the next frequency level of general English (see the third row of Table 3).

The last row of Table 6 presents the potential coverage of the ASWL for learners with knowledge of the most frequent 3,000 word-families. Learning 75 word-families from Level 4 ASWL would allow them to reach potential coverage of 95%-96%. This is particularly meaningful for learners studying in non-EMI courses. As shown in Table 3, they would need to learn 1,000 word-families from the 4th 1,000-word level of general English to reach 95% coverage of lectures in this context. Focusing on ASWL words instead, these learners need to study a much smaller number of words but can still reach 95% coverage. Learners studying in EMI courses or open access non-EMI courses may be able to understand lectures in these contexts thanks to their current vocabulary knowledge. However, studying only 75 extra words would allow them to achieve higher lexical coverage, which would help them to improve lecture comprehension.

Table 6. Potential coverage achieved by learners with different vocabulary levels with knowledge of the ASWL words

Learners' prior knowledge of general vocabulary	Number of ASWL word-families to learn	ASWL levels to learn	Potential coverage (%)		
			EMI	Non-EMI	Open access non-EMI
< 1,000 words	1,741	1-4	93.32	91.24	93.72
1,000 words	911	2-4	93.63	92.49	93.52
2,000 words	455	3-4	94.41	93.47	94.62
3,000 words	75	4	95.96	95.07	95.96

4. Discussion

4.1. Vocabulary in lectures in different contexts

The present study provides further insight into the nature of vocabulary in lectures in different contexts. First, it revealed that the difference in the lexical demand across contexts became more obvious as the degree of desired comprehension increased. If 'relatively high' comprehension

(90% coverage) was set as the goal, there was no difference in the vocabulary load of lectures in the three contexts. They all required knowledge of the most frequent 2,000 word-families. However, if ‘relatively high and stable’ comprehension (95%) was the goal, lectures in non-EMI courses would require a larger vocabulary size (4,000 word-families) than those in the other two contexts (3,000 word-families). If ‘very high’ comprehension (98% coverage) was set as the goal, learners would need to know 9,000 word-families for non-EMI courses, but only 8,000 word-families for open access non-EMI courses, and 7,000 word-families for EMI courses. The vocabulary sizes needed for comprehension of lectures in non-EMI courses and open access non-EMI courses were comparable to those found by Dang and Webb (2014), but the vocabulary sizes needed for comprehension of EMI courses were smaller. This finding is probably because the lectures and seminars in Dang and Webb’s study were from non-EMI courses.

Taken together, the findings of this study suggest that from the perspectives of vocabulary load, if ‘relatively high’ comprehension is the target, the demand of lectures in the three contexts is fairly similar. However, if higher levels of comprehension are desired, lectures in non-EMI courses are likely to be the most demanding, followed by those in open access non-EMI courses; lectures in EMI courses appeared to be the least demanding.

These findings could be the result of the difference in the English language proficiency of lecturers and students in the three contexts. Lectures in non-EMI courses are the most lexically demanding because these courses tend to have a larger proportion of lecturers and students who are English L1 speakers than courses from the other two contexts. Moreover, EAL students in non-EMI courses need to meet higher language entry requirements (Baker & Hüttner, 2017) and have greater opportunities to emerge themselves in English speaking environments. Therefore, they may have higher language proficiency levels than those from the other contexts.

The fact that open access non-EMI lectures were the second most demanding in terms of lexical coverage could be explained as follows. These lectures were also recorded in English-speaking countries, and thus may have a similar proportion of lecturers who are English L1 speakers as non-EMI courses. However, unlike non-EMI courses, students in open access non-EMI courses are more diverse, including both English L1 speakers and EAL users. Many of them are from low socioeconomic backgrounds and have poor levels of academic skills and low English language proficiency (Liyanagunawardena et al., 2013). Therefore, generally students in open

access non-EMI courses tend to have lower English language proficiency than those studying in non-EMI courses. To accommodate students' comprehension, lecturers in open access non-EMI courses may need to use simpler vocabulary than lecturers in non-EMI courses. Additionally, although most sessions in the open access non-EMI courses were recorded spontaneously in real classrooms, several sessions were pre-recorded lectures, whose scripts had been prepared in advance. The scripted nature of those lectures may give lecturers time to carefully select the vocabulary that facilitates learners' comprehension. This is suggested by the fact that words outside the most frequent 3,000 word-families accounted for a smaller coverage in open access non-EMI corpus (4.55%) than in the non-EMI corpus (5.12%).

Lectures in EMI courses were the least lexically demanding perhaps because lecturers in these courses use simpler language than those in the non-EMI and open access non-EMI courses. Most of the lecturers and students in EMI courses are likely to be EAL users who have limited exposure to English language outside classroom (Mauranen et al., 2016; Tamtam et al., 2012). Moreover, the language entry requirements of courses in EMI settings are likely to be lower than those in non-EMI settings (Baker & Hüttner, 2017). For these reasons, the English language proficiency of learners in EMI courses tends to be lower than those studying in EMI courses and open access non-EMI courses. As a result, lecturers need to adjust their lectures to accommodate students' language proficiency. Additionally, many EAL lecturers in courses in the EMI courses are not confident in their English language proficiency (Fortanet-Gómez, 2012) and may use simpler language.

4.2. Vocabulary load of lectures in comparison with other kinds of academic speech events

This study indicates that lectures are more demanding than labs and tutorials. Regardless of the contexts, this study found that the vocabulary sizes needed to reach 90%, 95%, and 98% coverage of academic lectures were generally higher than those needed for labs (1,000 word-families, 3,000 word-families and 7,000 word-families) and tutorials (1,000 word-families, 3,000 word-families and 4,000 word-families) as reported by Coxhead et al. (2017). This difference may be due to the nature of these speech events. First, lectures are events in which lecturers convey, elaborate, and evaluate subject content (Basturkmen, 2016; Deroey & Taverniers, 2011; Lynch, 2011). Meanwhile, labs and tutorials are events in which students discuss with their tutors and other students the problems related to the lecture content (Adolphs & Carter, 2013;

Basturkmen, 2016). Because of their purposes, lectures are denser in subject content and may contain a larger number of low-frequency words than other kinds of speech events, which explains why lectures have a heavier vocabulary load than the other speech events. Second, lectures tend to be more monologic than labs and tutorials (Basturkmen, 2016; Deroey & Taverniers, 2011). The main speakers of lectures are course instructors while in labs and tutorials, students are the main speakers and course instructors only act as facilitators. Therefore, the language in lectures is mainly the lecturers' language whereas the language in seminars, labs, and tutorials is mainly students'. This might be another reason why lectures are more lexically demanding than labs and tutorials.

4.3. The relative value of the ASWL for learners in different contexts

This study provides further insight into the value of the ASWL for learners in different contexts. All four levels of the ASWL benefit learners regardless of their vocabulary levels and learning context. Learners only need to study a small number of word-families but can achieve high lexical coverage, which helps to improve their comprehension of lectures. Moreover, despite the variation in frequency ranking of ASWL words in each context, all ASWL words occurred in the three subcorpora. However, the relative value of each ASWL level may vary according to the context. Levels 1-3 of the ASWL seem more beneficial for learners studying in open access non-EMI courses or EMI courses. Research has found that EAP learners planning to study in EMI courses have insufficient knowledge of the most frequent 2,000 word-families and even the most frequent 1000 word-families of general English after many years of formal English instruction (Akbarian, 2010; Dang, 2020b). Ideally, EAP programs should support this group of learners to learn the most frequent 3,000 word-families of general English to achieve reasonable comprehension of academic lectures. However, in many contexts, class time is limited, and the vocabulary growth rate of learners in EFL contexts is fairly slow (Webb & Chang, 2012). Therefore, focusing on the ASWL offers a shortcut to learners who have insufficient knowledge of the most frequent 2,000 word-families. Compared to studying items from the next 1,000 word levels of general English, studying the ASWL words that are beyond their current vocabulary levels would allow learners to learn a smaller number of words, but still achieve reasonable comprehension of academic lectures. Level 4 of the ASWL seems to be the most beneficial for learners studying non-EMI courses. To achieve reasonable comprehension of academic lectures, these learners would need to study 1,000 words from the 4th 1,000 words of general vocabulary.

However, if they learn only 75 words from Level 4 of the ASWL instead, they would still achieve reasonable comprehension. It should be noted that the discussion is based on the assumed prior vocabulary knowledge of the majority of learners in each context; there may be learners in non-EMI courses with insufficient knowledge of the most frequent 3,000 word-families, who would then benefit more from Levels 1-3. In contrast, some learners in open access non-EMI courses or EMI courses may have sufficient knowledge of the most frequent 3,000 word-families and may benefit more from Level 4.

5. Limitations and future research

This study has several limitations. First, because of the availability of resources, open access EMI courses were not examined, and the EMI corpus was much smaller than the other corpora and only sampled lectures recorded in Hong Kong and Finland. Therefore, the findings of this study are only exploratory. Second, this study is corpus-driven. Further research with a larger corpus of EMI lectures and using other methods (e.g., interviews, questionnaires, classroom observations with learners and teachers) would provide further insight into the lexical demands of lectures and the value of using the ASWL in EAP programs.

6. Pedagogical implications

This study provides several implications for EAP teachers. Teachers should aware that the lexical demand of lectures varies according to the degree of desired comprehension and contexts. If ‘relatively high’ comprehension of academic lectures (90% lexical coverage) is set as the goal, the amount of vocabulary that learners need to know is fairly similar in the three contexts. If ‘relatively high and stable’ comprehension (95% lexical coverage) is the goal, the number of words learners need for open access non-EMI courses and EMI courses are similar, but learners studying in non-EMI courses will need to know a larger number of words. If ‘very high comprehension’ is the target (98% lexical coverage), students planning to study in non-EMI courses would need to know more words than those planning to study in open access EMI courses, who in turn need to know a larger number of words than those planning to study in EMI courses.

Furthermore, this study suggests that the ASWL is a useful resource for EAP learners to enhance their comprehension of academic lectures. Teachers can use this list as a guide to set the vocabulary learning goals and design learning activities and materials for their students.

Although ASWL levels are useful for learners regardless of their contexts, Levels 1-3 seem to be more useful for those planning to study in EMI courses and open access non-EMI courses while Level 4 is especially useful for those studying in non-EMI courses. To decide which ASWL words learners with different vocabulary levels should focus on, teachers could follow Dang et al.'s (2017) learning sequences when selecting items from the ASWL. In this way, learners are introduced the most frequent items before the least frequent items. As a result, they can learn the lexical items that they are more likely to encounter in academic lectures first.

It is important to note that although the ASWL is presented as a list of word forms and does not disambiguate meanings, it does not mean that teachers should focus solely on directly teaching the forms of these lexical items through decontextualized methods. Instead, they should follow Nation's (2007) 'Four Strands' principle to organize language-focused learning, meaning-focused input, meaning-focused output, and fluency development activities so that students have opportunities to repeatedly encounter and use the target lexical items in various contexts. Some possible Four Strand activities for learning the ASWL words are presented below.

Language-focused learning activities focus on explicitly teaching the ASWL words so that learners attend to these words, which increases the likelihood that these items are learned. This strand is especially important for Levels 1-3 ASWL words. As these words are also common in general English, they may not be very salient to learners and will need explicit instruction (Dang, 2020b). To present the ASWL words in specialized contexts, teachers could implement the data-driven learning (DDL) approach (Johns, 1991). First, teachers can find transcripts of lectures in academic spoken corpora that are relevant to their specific contexts. Samples of lectures non-EMI courses can be found in the BASE⁴ and MICASE corpora. Samples of lectures from EMI courses can be found in the ELFA corpora. Samples of lectures from open access non-EMI courses can be found from open coursewares which are delivered by MIT and Yale University. Then, teachers can upload the transcripts to corpus analysis toolkits for lexical analysis such as Sketch Engine or Antconc and search for the target ASWL words. Next, they can provide students with these transcripts in which the target lexical items are highlighted and ask them to

discuss the meaning and use of these lexical items in the lectures. These language-focused learning activities would help students connect the forms, the general meanings, and the specific meanings of the ASWL words in academic lectures. Where students are comfortable with using technology in learning, teachers can instruct them how to use corpus-based tools such as Sketch Engine and Antconc to check the occurrences of ASWL words in lecture transcripts.

While language-focused learning activities help students to learn the ASWL words deliberately, meaning-focused input activities allow learners to realize how the ASWL words are used in different contexts and to consolidate knowledge of known items. Extensive viewing of lectures is a useful meaning-focused input activity for learning ASWL words (Dang et al., 2021). TED talks and subject-related television programs are also useful resources as they contain a substantial amount of specialized vocabulary that occurs in academic lectures (Dang, 2020a; Liu & Chen, 2019).

The primary focus of learning the ASWL is to develop learners' receptive knowledge of these lexical items so that they can improve their comprehension of academic lectures. However, Nation (2007) points out that a principled vocabulary learning program should allow learners to repeatedly encounter and use vocabulary in varied contexts. Therefore, meaning-focused output activities for learners to develop their productive knowledge of the ASWL words are also important. For example, teachers can organize activities in which students make presentations on specific topics in which they use the ASWL.

The last strand is fluency development. Activities in this strand do not focus on learning new lexical items but help learners to process and use known ASWL words faster. For instance, teachers can ask students to listen to the same lecture several times with the speech of the lecture increasing over the time. Another example is 4/3/2, in which students talk about the same topic to different audiences three times with the time of the talk reducing gradually from 4 to 3 and then 2 minutes.

7. Conclusion

Expanding on previous research, this study provides a comprehensive picture of vocabulary coverage in academic lectures in different contexts. It revealed that from the perspective of lexical coverage, lectures in non-EMI courses are likely to be the most demanding, followed by

lectures from open access non-EMI courses; lectures in EMI courses are the least demanding. This study also indicates that the ASWL is a useful shortcut for learners to improve their lecture comprehension. Moreover, although learners in all contexts can benefit from this list, those in EMI courses and open-access non-EMI courses would benefit more from Levels 1-3 while those in non-EMI courses would find Level 4 more useful.

Notes

¹ Some parts of these corpora were from the corpus used to develop Dang et al.'s ASWL, which consisted of materials from the BASE, MICASE, PICASE, LIBEL, HKCSE, open courses delivered by Stanford University, Yale University, Massachusetts Institute of Technology, University of California at Berkeley, and lectures at a university in New Zealand. However, new materials were also added. They were transcripts of the lectures from the mentioned sources (but not included in the ASWL corpus) as well as those from the English as a Lingua Franca in Academic Settings (ELFA) corpus.

² The raw transcripts are available at www.coventry.ac.uk/base. However, the BASE team has made it open access via Sketch Engine.

References

- Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. Routledge.
- Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38(3), 391–401.
- Baker, H., & Hüttner, J. (2017). English and more: A multisite study of roles and conceptualisations of language in English medium multilingual universities from Europe to Asia. *Journal of Multilingual and Multicultural Development*, 38(6), 501–516.
- Basturkmen, H. (2016). Dialogic interaction. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 152–164). Routledge.

- Becher, T. (1989). *Academic tribes and territories*. The Society for Research into Higher Education and Open University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–362.
- Coxhead, A. (2016). Reflecting on Coxhead (2000), “A New Academic Word List”. *TESOL Quarterly*, 50(1), 181–185.
- Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 97–110). Routledge.
- Coxhead, A., & Dang, T. N. Y. (2019). Vocabulary in university tutorials and laboratories. In K. Hyland & L. Wong (Eds.), *Specialised English: New directions in ESP and EAP research and practice* (pp. 120–134). Routledge.
- Coxhead, A., Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30, 66–78.
- Dang, T. N. Y. (2018a). A Hard Science Spoken Word List. *ITL-International Journal of Applied Linguistics*, 169(1), 44–71.
- Dang, T. N. Y. (2018b). The nature of vocabulary in academic speech of hard and soft sciences. *English for Specific Purposes*, 51, 69–83.

- Dang, T. N. Y. (2020a). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly*, *54*(2), 436–459.
- Dang, T. N. Y. (2020b). High-frequency words in academic spoken English: Corpora and learners. *ELT Journal*, *74*(2), 146–155.
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997.
- Dang, T. N. Y., Lu, C., & Webb, S. (2021). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition, Open Access*, 1–29. <https://doi.org/10.1017/S0272263121000474>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76.
- Dearden, J. (2014). English as a medium of instruction: A growing global phenomenon. *British Council*.
- Deroey, K. L. B., & Taverniers, M. (2011). A corpus-based study of lecture functions. *Moderna Språk*, *2*, 2–22.
- Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 1–13.
- Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, *6*, 3–17.
- Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, *30*, 198–208.

- Fortanet-Gómez, I. (2012). Academics' beliefs about language use and proficiency in Spanish multilingual higher education. *AILA Review*, 21(1), 48–63.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range: A program for the analysis of vocabulary in texts*. <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Hyland, K. (2016). General and specific EAP. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 17–29). Routledge.
- Johns, T. (1991). Should you be persuaded—two samples of data-driven learning materials. In T. Johns & P. King (Eds.) *Classroom concordancing* (pp. 1-6). ELR Journal Vol. 4. Birmingham: Centre for English Language Studies, University of Birmingham.
- Liu, C. Y., & Chen, H.-H.-J. (2019). Academic spoken vocabulary in TED Talks: Implications for academic listening. *English Teaching & Learning*, 43, 353–368.
- Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *International Review of Research in Open and Distance Learning*, 14(3), 202–227.
- Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10(2), 79–88.
- Mauranen, A., Hynninen, N., & Ranta, E. (2016). English as the academic lingua franca. In K. Hyland & P. Shaw (Eds.) *The Routledge handbook of English for Academic Purposes* (pp. 44–55). Routledge.
- Mulligan, D., & Kirkpatrick, A. (2000). How much do they understand? Lectures, students and comprehension. *Higher Education Research & Development*, 19, 311–335.

- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 1–12.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*.
<http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, 52, 188–200.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Tamtam, A. G., Gallagher, F., Olabi, A. G., & Naher, S. (2012). A comparative study of the implementation of EMI in Europe, Asia and Africa. *Procedia - Social and Behavioral Sciences*, 47, 1417–1425.
- Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253–270). Peter Lang.
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479.

- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56–89.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258.
- Wallace, M. (2020). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*.
<https://doi.org/10.1111/lang.12424>
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28, 170–182.
- Watson-Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, 45, 31–39.
- Webb, S., & Chang, A. C.-S. (2012). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651–675.