International Journal of TESOL Studies (2023)
Vol. 5 (3) 126-142 https://doi.org/10.58304/ijts.20230310

Article

A Corpus-Based Mechanical Engineering Academic Word List

Le Chang
Shenyang Ligong University, China

Received: 29 May 2022; Accepted: 22 March 2023; Published: 10 June 2023

Abstract
Most learners of a second or foreign language fall short of necessary academic vocabulary, and the problem has become increasingly evident in the context of technical fields such as mechanical engineering. This paper reports on the development of a corpus-based academic word list for mechanical engineering graduate students from a total of 120 academic articles collected from four internationally recognized journals. The created corpus called Mechanical Engineering Academic Corpus (MEAC) has 755,794 words. With Coxhead's word selection criteria and Range software programme for lexical analysis, a 398-word/289-word-family Mechanical Engineering Academic Word List (MEAWL) was constructed. Comparisons show that MEAWL is only approximately half of Coxhead's Academic Word List (AWL) in size (289 word families vs.570 word families) but has a higher coverage in MEAC than AWL (12.82% vs.11.95%), and therefore may be more suitable for mechanical engineering students, who will spend much less time on vocabulary learning. The practical value of MEAWL such as its pedagogical implication in material development is also discussed.

Keywords
Academic word list, corpus, mechanical engineering, coverage

1 Introduction

Vocabulary is an indispensable component of any language, and it has been widely accepted that language competence depends primarily on adequate lexical knowledge. According to Nation (2013), words in English academic writing can be categorized into high-frequency words, academic words, technical words and low-frequency words. The most noticeable lexical difficulty for English as a foreign language (EFL) learners exists neither in high-frequency words which occur in almost all types of texts, nor in technical words for they are central to students' specialized areas and are often encountered in the lectures, but in between, i.e., academic words.

1.1 Academic words

Academic words are described as being "formal, context-independent with a high frequency and/or

Email: gmwa@163.com

range of occurrence across scientific disciplines" (Ferrell, 1990, p. 11). They perform discourse functions in raising research questions, reviewing related literature, describing methods, discussing findings, drawing conclusions, etc. To language users, academic lexical knowledge is such a key indicator of their academic competence and important factor affecting their academic reading and writing that a growing body of evidence (e.g., Laufer and Nation, 1999; Gardner & Davies, 2014) argues that without adequate academic vocabulary, English for Academic Purposes (EAP) learners cannot deal with academic materials effectively.

However, academic vocabulary cannot be encountered in a text as often as high-frequency general words. Studies (e.g., Coxhead & Byrd, 2007) have found that academic words appear very minimally in daily reading. To be specific, academic words only cover 1.4% and 4.5% of the words in novels and news respectively, and can hardly be acquired without intentional study. It has become evidently imperative for EAP teachers to realize and emphasize the importance of academic words and provide learners with focused academic vocabulary instruction.

One of the most efficient ways to gaining academic words is via an appropriate corpus. A corpus is defined as a body of text representative of a particular variety of language stored on a computer, and corpora have grown into an indispensable methodology to study natural language on examples of "real life" language use (McEnery & Wilson, 2001). In particular, when a collection of the occurrences of a word-form (i.e., a concordance) is computerized, language facts become more explicit along with the production of vocabulary lists and lexical syllabuses for EAP courses.

1.2 General academic word lists

To identify particular high-frequency words worth for students learning, researchers have compiled various academic word lists, either general or specific. General word lists concern common academic words of various disciplines and thus are also labeled as discipline-crossing (e.g., Veenstra & Sato, 2018), whereas specific word lists, with a focus on academic words of a specific discipline, are also referred to as discipline-based lexical repertoires (Hyland & Tse, 2007) or field-specific academic word lists (Martinez et al., 2009).

The earliest major general word list, which is till influential to date, is the General Service List (GSL) of English words created by West (1953) on the basis of a 5-million-word corpus of written English. GSL aimed to identify the 2,000 most frequent word families (a word family refers to a stem plus all its related reflections and derivations) useful for English learners and has fundamentally defined the notion of core vocabulary in English. However, since language changes overtime, a New GSL containing updated words has been developed by Brezina and Gablasova (2015) to suit current EFL courses, though yet to replace the more established 1953 one.

One of the most extensively cited and applied general word lists is Coxhead's (2000) Academic Word List (AWL), a 570-word-family word list based on her 3.5-million-word Academic Corpus from reading materials of 28 different subject areas in four disciplines (arts, commerce, law and science). To be included in AWL, words had to meet three criteria, which many subsequent studies have followed or discussed: (1) specialized occurrence — GSL was excluded due to the purpose of AWL being for academic but not general vocabulary; (2) frequency — the minimum occurrence of each word in AWL was fixed on 100 times in the corpus; and (3) range — to avoid the situation that "a word count based mainly on frequency would have been biased by longer texts and topic-related words"(Coxhead, 2000, p. 221), all the words in AWL had to occur at least 10 times in each of the four disciplines and at least 14 of the 28 subject areas to represent a variety of academic disciplines.

In an attempt to prove the validity of an academic word list in its application to academic learning and research, Hu and Nation (2000) used the notion of text coverage, i.e., the percentage of running words in the text known by the readers. Coxhead (2000) found AWL and GSL respectively covered 10.1% and 76.1% of the running words in her corpus, resulting a high combined coverage of over 86%. Subsequently, other researchers (e.g., Hyland & Tse, 2007; Valipouri & Nassaji, 2013) examined the generalized validity of AWL in different disciplines and found it accounted for a consistent coverage of about 10% in various academic texts. Thus, according to Coxhead and Nation's (2001) paradigm of English words, after mastering AWL (10%), in combination with the high-frequency general words in GSL (76.1%), technical words (5%), and the already known proper nouns and abbreviations (5%), learners could understand 95% of the total words in their reading, a lexical coverage threshold which guarantees comprehension of academic texts (Schmitt et al., 2017).

Nevertheless, some researchers consider AWL to be too general and question its effectiveness in developing academic vocabulary for EAP courses. Hyland & Tse (2007) brought forward the first independent evaluation of AWL and discovered in their own 3.3-million-word corpus that despite the 10% coverage of AWL, the distribution was quite unbalanced across the disciplines — a 16% coverage for computer sciences and a 6.2% coverage for biology. Similarly, Wu and Wang (2007) found AWL only covered 9.3% in Jiaotong Daxue English of Science and Technology Corpus, and out of the 570 word families in AWL, Martinez, Beck and Panza (2009) only counted 92 which frequently appeared in agriculture research articles. Given that, Liu and Han (2015) suggested that AWL should not be used in any discipline without adaptation, and other researchers (e.g., Hyland & Tse, 2007; Liu & Lei, 2020) even argued that language teachers should approach AWL with caution and endeavour to develop more restricted, discipline-based word lists to meet the specific academic needs of learners. Engineering students, for example, would benefit more from learning words directly associated with their subject matters such as *load*, *stress*, or *machining*, instead of AWL words like *legal* or *finance*, which are more intended for learners of law or finance disciplines.

To compile a word list from a corpus larger and more recent than the one used for AWL, Gardner and Davies (2014), created a new Academic Vocabulary List (AVL). The academic corpus comprised 120 million words of academic texts as part of the 425-million-word Corpus of Contemporary American English (COCA) which was composed of nine academic disciplines such as education, history, science and technology, etc. AVL is known for the thorough methodological procedures employed in its creation: ration, range, dispersion, and discipline measure. The 3015-word AVL was found to cover 13.8% of the academic materials in COCA and have nearly twice the coverage as AWL when the top 570 word families of AVL were compared with those of AWL. Therefore, AVL is believed to be "the most current, accurate, and comprehensive list" (Gardner and Davies, 2014, p. 325). However, AVL is not without limitations. For instance, Durrant (2016) found that even though 427 words out of the total list are frequent across 90% of disciplines, it is a considerably small scale in contrast to the entire AVL. To be more specific, about half of the words in the list are hardly used.

1.3 Engineering academic word lists

Nation (2016) argues that each academic discipline has its own specific vocabulary closely associated with the content involved in that area. The aforementioned disadvantage of general academic word lists has stimulated some language teachers and researchers to focus on academic words used in single disciplines. Academic word lists in engineering disciplines (Table 1) can be classified as general engineering academic word lists which concern academic vocabulary across engineering subjects and specific engineering academic word lists which are more subject-specific in nature, i.e., word lists specifically developed for a certain engineering subject.

Table 1

Engineering Academic Word Lists

Researcher	Corpus	Word List and its Size		Coverage
Mudraya (2006)	Student Engineering Corpus (SEC); 13 coursebooks; 2 million words	Student Engineering Word List (SEWL)	1260 word families / 8850 words	Not provided
Ward (2009)	Engineering Corpus (EC); 25 coursebooks (chemical, civil, electrical, industrial and mechanical); 271,000 words	Basic Engineering List (BEL)	299 words	16.4%
Hsu (2014)	Engineering Textbook Corpus; 100 coursebooks (automation, chemical, civil, electrical, etc.); 4,570,000 words	Engineering English Word List (EEWL)	729 word families	14.3%
Veenstra & Sato (2018)	Science Textbook Corpus; 12 coursebooks (biology, chemistry, physics, and engineering); 704,237 words	Science Textbook Word List (STWL)	309 word families	13.4%
Jiang (2010)	Petroleum English Corpus; 5 coursebooks (petroleum drilling, storage and transportation, exploration, petrochemical and development English); 180,000 words	Petroleum Academic Word List (PAWL)	498 word families	11.3%
Zhao (2015)	Maritime English Corpus (MEC); coursebooks, articles, magazines, documents, and conventions; 11 topics; 3 million words	Maritime English Academic Word List (MEAWL)	438 word families / 641 words	11.78%

The first documented general engineering academic word list, namely the Student Engineering Word List (SEWL) was built by Mudraya (2006) based on a Student Engineering Corpus compiled from 13 textbooks. SEWL did not distinguish basic words (GSL) and non-basic words and therefore resulted in a lengthy word list of 8850 words (1260 word families), which may overburden learners. In contrast, Ward (2009) completed from 25 engineering coursebooks a much shorter 299-word Basic Engineering List (BEL), a list that could be undertaken at an earlier phase of college for his low-proficiency undergraduate students in Thailand. Ward found BEL had 16.4% coverage in his corpus of coursebooks and emphasized that teachers should not only focus on individual words but also their lexico-grammatical environments.

Hsu (2014) compiled 100 college coursebooks across 20 engineering subjects into a 4.57-million-word Engineering Textbook Corpus and produced a 729-word-family Engineering English Word List (EEWL), which was reported to cover 14.3% of the words in the coursebooks. Hsu hence claimed that, on the basis of GSL, learners could achieve effective reading comprehension of all fields of engineering coursebooks after 1-2-semester study of EEWL. Likewise, to assist their 1st-and-2nd-year university students in Japan, Veenstra and Sato (2018) combined 12 science and engineering coursebooks and constructed a corpus-based Science Textbook Word List (STWL), a general science and engineering academic vocabulary list which contained 309 word families beyond the GSL and the text coverage of STWL in the 12 coursebooks was found to be 13.4%.

Ward (2009) found uneven distributions of Mudraya's (2006) common engineering words in the five sub-disciplines of his Engineering Corpus. He listed 10 most extreme examples (Table 2) to support his claim that engineering disciplines are lexically idiosyncratic. As is evidently displayed in the table, none of these common engineering words are important for industrial engineering, few for electrical engineering and only a few for chemical engineering. Following Ward's argument that "time spent teaching *fluid* to civil engineering students would be better spent on words like *load*" (p. 173), some researchers (e.g., Zhao, 2015) have attached focus on developing academic word lists to help learners in some specific engineering disciplines such as petroleum and maritime.

Table 2

Distribution of 10 Common Engineering Words in Engineering Corpus (EC) (taken from Ward, 2009, p.173)

Word family	Total frequency in EC	Chemical Engineering (EC)	Civil Engineering (EC)	Electrical Engineering (EC)	Industrial Engineering (EC)	Mechanical Engineering (EC)
pressure	376	145	103	3	6	119
velocity	239	82	4	11	1	141
stress	507	6	253	0	1	247
load	537	7	265	138	1	126
body	87	3	12	12	3	57
moment	257	0	208	4	2	43
fluid	134	85	2	5	1	41
beam	404	1	222	1	5	175
shear	337	4	195	1	1	136
equilibrium	192	72	53	6	0	61

In finding that AWL only covered 8.1% words in her 180,000-word Petroleum English Corpus and 12% AWL words never occurred in it, Jiang (2010) constructed a 498 word-family Petroleum Academic Word List, with 72 word families fewer than AWL but 3.2% higher coverage. In the same way, Zhao (2015) compiled some English coursebooks, articles, magazines, documents and conventions into a 3-million-word Maritime English Corpus, from which he created a Maritime English Academic Word List composed of 438 word families or 641 words. The word list reported a high coverage of 11.78% in his corpus, 3.04% higher than AWL.

With no doubt, each of the aforementioned word lists can best suit its intended learners, either in a general academic manner or for a specifically engineering purpose. But to date, a mechanical engineering academic corpus has not been reported, let alone any corpus-based academic word list in the sub-discipline, though one would be of great value. With the increasing needs to communicate academically and publish internationally, a mechanical engineering academic word list will help students, researchers and practitioners of the field in this endeavour.

1.4 The Present study

The present study aims to specifically address the needs of mechanical engineering graduate students. The subjects, at a university in Northeastern China, had studied EFL for at least 10 years (6 years in middle schools and 4 years in undergraduate period), and were supposed to have already had a fair mastery of the basic general vocabulary in GSL. Doing graduate studies, they were expected to do at least part of reading and writing in English to achieve the ultimate goal of understanding English academic materials in their field and submitting research proposals and articles internationally. Therefore, it is imperative that they shift their limited amount of time from studying general academic vocabulary to learning mechanical engineering academic words as required by their specific field.

To this end, two main aims of the study have emerged: to create for mechanical engineering graduate students a reliable academic word list from a corpus complied by academic articles in the field; and to compare the created word list with AWL to measure the overlap and difference so as to determine which word list could better suit the needs of the intended learners. In accordance with these aims, the research seeks to address two questions:

1. Beyond GSL, what vocabulary items occur with reasonable frequency and range in a mechanical engineering research article corpus so that they can form a Mechanical Engineering Academic Word List (MEAWL)?
2. How do the words in MEAWL compare to those in AWL?

2 Methodology

The present study was designed to first construct a mechanical engineering academic corpus and then compile a mechanical engineering academic word list.

2.1 The corpus

Because the ability to read and write academic articles is the fundamental concerns for most graduate students, the corpus was designed to be composed by a single genre, i.e., research articles. First, a mechanical engineering professor was invited to divide the discipline into its subject areas, which turned out to be mechanical manufacturing and its automation, mechanical design and theory, and mechanical and electronic engineering. To fulfil the representativeness of the corpus, the professor was requested to recommend some internationally regarded academic journals for these subject areas, and 4 journals — *International Journal of Machine Tools and Manufacture, Journal of Microelectromechanical Systems, Journal of Mechanical Design (Transactions of the ASME)* and *Mechatronics* were determined as the source journals, covering all three subject areas in the profession's particular context.

Table 3

Composition of MEAC

Subject Area	Journal	Article Number	Number of the Running Words
Mechanical Manufacturing and Its Automation	International Journal of Machine Tools and Manufacture	40	227,451
Mechanical Design and Theory	Journal of Mechanical Design (Transactions of the ASME)	40	309,393
Mechanical and Electronic Engineering	Journal of Microelectromechanical Systems;	28	218,950
	Mechatronics	12	
Total 3	4	120	755,794

In compiling the corpus, 120 articles published between 2015 and 2020 in the conventionalized IMRD (Introduction-Method-Results-Discussion) format as Swales (1990) suggested were downloaded and then saved into different text files in accordance with the varied subject areas. To reach uniformity, each area included 40 articles (as shown in Table 3), and the length of the articles were balanced as much as possible. Next, some unreadable components for the computer software programme Range (see more description of Range in next section) and unrelated factors to lexical analysis, such as tables, graphs, equations, footnotes, appendices, references, acknowledgments, etc. were removed in the corpus cleaning process. Finally, a 120-text, 755,794-word Mechanical Engineering Academic Corpus (MEAC) was constructed.

2.2 Range software

Range, a computer software program developed by vocabulary researchers Nation and Heatley (www.vuw.ac.nz/lals/staff/paul-nation) was used in the study for word list creation and lexical analysis. Range can sort words by their frequency (times of a word's occurrence), range (areas of a word's occurrence in) as well as their alphabetical order, making it a useful tool to create a word list and then examine the frequency and range of a word in a given text. Some ready-made base lists, i.e., GSL and AWL are used as files in Range. To be more specific, GSL is divided into two files — the first and second most common 1,000 words, and the words in AWL comprise the third file. Though, words outside GSL and AWL are also provided in Range. The purposes of using Range in the study were three-fold: to compile a Mechanical Engineering Academic Word List (MEAWL); to analyze the coverage of MEAWL and GSL in MEAC; and to compare word families in MEAWL and AWL.

2.3 Word selection criteria

The present study basically followed the criteria of Coxhead (2000) in word selection. First, concerning the specialized occurrence, due to the fact that the intended learners in the study were graduate students, all words in the list were decided to be outside the most common 2,000 basic words and GSL was thus excluded. Secondly, with respect to range, to represent a variety of subject areas within the discipline, it was decided that words must appear in at least two of the three subject areas of mechanical engineering. Last, in terms of frequency, words in Coxhead's corpus of 3.5 million words had to occur at least 100 times in order to be selected. Hence, to ensure the text coverage in the study, the minimum of word occurrence in the 755,794-word MEAC was calculated and fixed on 22 times.

In brief, for words to be included in the word list of the study, the following criteria must be satisfied: (1) Specialized occurrence: outside GSL; (2) Range: at least two subject areas of mechanical engineering; and (3) Frequency: minimum occurrence of 22 times. With these three criteria fulfilled and after the final confirmation of experts in the discipline, a 398-word Mechanical Engineering Academic Word List (MEAWL) was finally created (Appendix 1).

3 Results and Discussion

In this section, results and discussion are presented in the order of the two research questions under study.

3.1 MEAWL and its coverage

To address the vocabulary items in MEAWL, Table 4 lists the most common 26 words, i.e., words occurring at least 80 times in all the three subject areas in MEAC. Apparently, these words are mostly nouns like *process*, *mechanism*, *simulation*, etc. reflecting the unique nature of mechanical engineering discipline, and verbs in their passive forms like *defined*, *obtained*, and *generated*, in line with Hyland's (2008) claim that to show their objectiveness scientific articles usually use passive sentences, so notional verbs generally appear in the form of past participles. In addition, words as *method*, *data*, *section*, *analysis*, and *reference* can reflect the component genre of MEAC from which MEAWL was extracted, i.e., research articles.

Regarding coverage, Table 5 shows that MEAWL accounts for 11.01% of the 755,794 running words in the whole MEAC. In other words, the lexical items from MEAWL appear averagely 11 times in each 100 words in MEAC.

A more detailed observation only finds minimal coverage difference (less than 2%) across different mechanical engineering subject areas — the area of mechanical manufacturing and its automation has similar coverage to that of mechanical and electronic engineering, i.e., 11.82% and 11.52% respectively, and the coverage of MEAWL in the area of mechanical design and theory is slightly lower (9.91%).

Table 4
The Most Common 26 Words in MEAWL

No.	Word	Frequency	No.	Word	Frequency
1	design	2178	14	similar	474
2	process	1557	15	range	431
3	method	906	16	defined	420
4	function	836	17	significant	368
5	data	776	18	area	366
6	parameters	709	19	structure	366
7	error	654	20	required	357
8	approach	628	21	accuracy	343
9	section	604	22	obtained	327
10	analysis	589	23	reference	318
11	contact	569	24	mechanism	310
12	methods	503	25	generated	293
13	maximum	476	26	simulation	255

Table 5 also shows that GSL covers 69.23% of the entire MEAC, an obvious distinction from Coxhead's (2000) finding of 76.1%. This indicates that MEAC contains a lower percentage of the 2,000 basic words and is more academic or technical in nature. A combination of GSL and MEAWL reaches a total coverage of 80.24%, suggesting that, with a foundation of the general basic words, mechanical engineering graduates may obtain sufficient lexical knowledge for academic articles reading in their discipline as long as they have a good command of the 398 words in MEAWL. When this coverage is supplemented by proper nouns, technical terms and their abbreviations, the learners are apt to meet the minimum requirement for effective academic reading comprehension in their own subjects.

Table 5

Lexical Coverage of MEAWL and GSL in MEAC

Subject area	Coverage of MEAWL	Coverage of GSL	Total Coverage
Mechanical Manufacturing and Its Automation	11.82%	71.17%	82.99%
Mechanical Design and Theory	9.91%	67.65%	77.56%
Mechanical and Electronic Engineering	11.52%	69.15%	80.67%
Whole Corpus (MEAC)	11.01%	69.23%	80.24%

Meanwhile, it is also noticeable that the total coverage varies across subject areas, with a maximum difference of over 5% (77.56% vs. 82.99%), indicating that there are more words beyond GSL and MEAWL for students to learn in the area of mechanical design and theory.

3.2 Comparison of MEAWL and AWL

The second research question attempts to find out the overlap and difference between MEAWL and AWL, so as to probe the usefulness of MEAWL. It should be noted that word family was used as the unit due to the fact that AWL was presented and analyzed in this manner.

As displayed in Table 6, after being input as base words in the Range software, the 570-word-family AWL accounts for 11.95% in MEAC, supporting Coxhead & Byrd's (2007) claim that AWL "covers approximately 10% of any academic text" (p. 132). Whereas MEAWL, reduced to 289 word families after grouping, achieves a coverage of 12.82% of the running words in MEAC, about 1% higher than that of AWL. The higher coverage of MEAWL is encouraging given that MEAWL is only about half size of AWL (289 vs.570 word families). In other words, mechanical engineering graduates may only need to learn 289 word families (or 398 words) instead of 570 word families to get an even higher lexical comprehension of the academic articles they intend to read.

Table 6

Word Families in MEAWL and AWL

Word List	Word Families	Coverage in MEAC	Word Families in Common	Overlapping Percentage
AWL	570	11.95%	151	26.5%
MEAWL	289	12.82%	151	52.2%

MEAWL's higher coverage with fewer vocabulary items can be explained by the fact that MEAWL was generated from MEC, a corpus solely compiled by mechanical engineering academic articles. Derived from such a more discipline focused corpus, MEAWL will certainly be of more efficiency, since it can save considerably much of the learners' time and relieve them from a great deal of vocabulary burden.

Table 6 also shows that MEAWL shares 151 word families with AWL (52.2% of MEAWL). This overlap can be explained by the fact that there are quite a number of "lexical items that occur frequently and uniformly across a wide range of academic materials" (Coxhead, 2000, p. 218). In other words, word families such as *analysis, factor,* and *significant* are general in academic texts across various disciplines.

On the other hand, the finding also indicates that beyond AWL there exist a good number of mechanical-engineering-specific academic words. To be more specific, there are 138 (289-151=138) word families in MEAWL for students of the discipline to learn, a finding revealing the shortage of AWL and proving the more usefulness of MEAWL over AWL for mechanical engineering graduate students.

Regarding the most used lexical items, after sublist 1 of MEAWL (i.e., the most frequently appeared 60 word families as presented in Table 7) was compared with its counterpart in AWL (Appendix 2), only 19 word families coincided (e.g., approach, data, function, similar, etc. which are underlined in Table 7) indicating that a majority of the most frequent academic words are not used similarly across disciplines. Words in sublist 1 of MEAWL such as *algorithm, chip, fabrication,* and *machining* cannot be found even in the entire AWL, not to mention in its sublist 1. The finding displays the mechanical engineering nature of MEAWL and reconfirms that AWL as a general word list should not be used in any specific discipline without alteration.

Table 7
Sublist 1 in MEAWL (in alphabetical order)

accuracy	design	identify	require
algorithm	device	layer	research
analyse	dynamic	linear	section
approach	element	machining	sensor
area	energy	maximum	significant
assembly	equation	mechanism	similar
axis	error	method	simulation
chip	evaluate	mode	specific
complexity	fabrication	obtain	stability
component	factor	optimization	structure
constant	feature	parameter	technique
contact	final	prediction	thermal
create	function	process	variation
data	generate	radius	vector
define	geometry	range	velocity

In agreement with the finding of Martinez et al. (2009), words in sublist 1 of AWL like *contract, export, legislate,* and *policy*, which reflect the main components of AWL being in commerce and law, never exist in MEAWL and thus are of little help if at all in reading and writing mechanical engineering academic articles. Therefore, using a generalized word list like AWL in its pre-determined sequence would possibly put engineering students at the risk of being exposed to unnecessary lexical items.

4 Conclusions

The first rationale for the study was to create a discipline-specific (mechanical engineering) and genre-specific (research articles) word list to meet the learning needs of graduate students in academic article reading and writing. The study brings forth a 398-word/289-word-family mechanical engineering academic word list (MEAWL) that accounts for 12.82% coverage of the whole MEAC, supporting the idea of developing field-specific vocabulary list from the target genres and texts that students need to read and write in their own academic discipline (Hyland & Tse, 2007; Martinez et al., 2009). In combination with GSL, MEAWL covers about 82% of the running words in MEAC. It may thus be concluded that with their prior basic vocabulary knowledge supplemented by MEAWL, mechanical engineering graduates are more probable to be sufficiently supported to comprehend research articles in their discipline.

The study also attempted to compare the coverage of MEAWL and AWL in MEAC, finding that with only a half size word list, MEAWL outperforms AWL in coverage. This has confirmed the assumption that a specialized word list rather than a general one can better facilitate discipline-specific students. To make it specific, with MEAWL mechanical engineering graduate students can follow an academic word list with considerably fewer lexical items than AWL but still understand the materials they are expected to read for their academic studies. Equally important is that from the teachers' perspective, they can spend valuable classroom time teaching a shorter list of words to yield greater benefit. Nevertheless, it must be emphasized that mechanical engineering students should develop general English proficiency before the competence of the specific vocabulary in MEAWL, for without a solid foundation, it would be difficult for them to be able to use the specific vocabulary competently, especially in academic writing.

5 Limitations

It should be acknowledged that the study has some limitations, with the most noticeable one being the small size of the corpus (about 0.75 million), a common characteristic of most discipline-specific corpora, leaving the result of the study, though useful, not the sole word list and certainly open to modification. Besides, following Coxhead's criteria, this study excluded the first 2000 basic words (i.e., GSL) in constructing the word list, a cut-off point future research should attach more consideration to, especially when the target users are at a disadvantaged level. Undoubtedly, a different academic word list will be generated when the threshold is settled at a lower level, for instance, the 1000 most frequent words. In addition, with MEAWL being a specialized word list tailored for mechanical engineering graduates, the study did not attempt to distinguish technical terms and sub-technical words, but counted them all academic words. However, it is suggested that future work should be cautious in examining whether the included words are indeed technical or non-technical/sub-technical, especially when the target users of the word list are at lower grades and technical terms should be learned through the subject specialist teachers.

6 Implications

A discipline-specific word list can "provide a useful guide for teachers to help them decide which vocabulary to focus on" (Nation, 2013, p. 258) and therefore, address the academic needs of language learners more directly than any general word list. The first practical value of MEAWL, due to its high frequency and range, can be discovered in its pedagogical implication as a guide for EAP teachers in designing coursebooks for reading and writing based on academic vocabulary. Besides, EAP teachers can also be inspired in material syllabus development by building MEAWL into a deliberately designed vocabulary learning program helping students with the target lexical items. Webb and Chang (2012) suggested that a goal of 400 English words a year could be achievable. With the objectives of most English courses being more than vocabulary learning, it is advised that teachers start from the top of the 398-word list, and classroom time be assigned to teaching reading and writing by following a principled approach, for instance, Nation's (2007) four strands to help students practise the words through input-, output-, and language-focused activities. Additionally, MEAWL can also be used as a reference list for students to learn by themselves. In the meantime, teachers who are well informed by MEAWL as a manageable resource of concordancers can have direct access to the target words via abundant authentic examples in the corpus. From a variety of contexts in MEAC, teachers can complement the drawback of a word list in being decontextualized. By bringing vocabulary instruction up to the levels of collocations or lexical chunks, teachers can help students understand word partnerships in the specific domain of mechanical engineering.

Appendix 1

Mechanical Engineering Academic Word List

No.	Word	Frequency	No.	Word	Frequency	No.	Word	Frequency
1	design	2178	2	process	1557	3	method	906
4	function	836	5	data	776	6	parameters	709
7	error	654	8	approach	628	9	section	604
10	analysis	589	11	energy	585	12	thermal	570
13	contact	569	14	layer	514	15	methods	503
16	maximum	476	17	errors	474	18	similar	474
19	machining	473	20	device	464	21	axis	446
22	sensors	444	23	constant	435	24	velocity	435
25	linear	432	26	geometry	431	27	range	431
28	defined	420	29	components	418	30	designs	418
31	complexity	413	32	sensor	412	33	assembly	410
34	chip	408	35	algorithm	398	36	optimization	398
37	stress	388	38	voltage	381	39	laser	376
40	feed	373	41	significant	368	42	area	366
43	structure	366	44	research	360	45	required	357
46	stability	350	47	impact	349	48	accuracy	343
49	volume	342	50	component	339	51	density	337
52	input	337	53	devices	329	54	features	329
55	obtained	327	56	response	326	57	torque	325
58	reference	318	59	profile	317	60	mechanism	310
61	objective	307	62	conventional	306	63	processing	303
64	vector	303	65	factor	301	66	specific	301
67	equations	299	68	matrix	296	69	dynamic	295
70	constraints	293	71	generated	293	72	parameter	292
73	ratio	292	74	equation	289	75	phase	289
76	final	282	77	complex	278	78	mode	276
79	damping	275	80	acceleration	270	81	distribution	269
82	fabrication	269	83	initial	266	84	elements	265
85	output	265	86	technique	261	87	hence	259
88	substrate	259	89	radius	258	90	diameter	257
91	orientation	255	92	simulation	255	93	previous	253
94	processes	252	95	factors	251	96	structures	250
97	technology	248	98	displacement	243	99	strategy	243

100	mechanisms	239	101	significantly	239	102	techniques	238
103	designed	237	104	selected	237	105	curvature	235
106	functions	235	107	machined	232	108	prediction	232
109	element	231	110	corresponding	230	111	individual	229
112	geometric	227	113	deformation	226	114	demonstrated	226
115	minimum	225	116	axial	223	117	feasible	222
118	physical	218	119	variation	217	120	feedback	213
121	overall	212	122	location	211	123	lateral	210
124	task	209	125	region	208	126	algorithms	207
127	negative	207	128	achieved	205	129	electrode	205
130	coupling	204	131	identified	204	132	approaches	203
133	potential	201	134	frequencies	200	135	sensing	199
136	trajectory	199	137	transition	198	138	create	197
139	loop	196	140	calibration	195	141	rotation	193
142	conducted	189	143	index	188	144	nozzle	186
145	achieve	184	146	source	183	147	available	182
148	dimensions	181	149	normal	181	150	scan	180
151	accurate	179	152	strain	175	153	positive	173
154	coefficients	172	155	magnitude	171	156	investigated	170
157	furthermore	169	158	dynamics	168	159	predicted	168
160	angular	167	161	evaluated	167	162	fabricated	166
163	approximately	165	164	assumed	165	165	modes	165
166	segment	165	167	requires	164	168	deviation	163
169	finally	161	170	fluid	160	171	layers	160
172	removal	160	173	require	159	174	created	157
175	identify	157	176	selection	157	177	target	157
178	jet	156	179	vacuum	154	180	derived	153
181	generate	153	182	lens	150	183	segments	150
184	estimated	149	185	optimal	149	186	parallel	148
187	requirements	148	188	transfer	148	189	environment	147
190	obtain	147	191	predict	147	192	affect	145
193	height	145	194	radial	145	195	shear	145
196	software	145	197	coordinate	144	198	technologies	143
199	concept	140	200	configuration	139	201	flexible	138
202	proportional	138	203	scanning	138	204	interface	137
205	demonstrate	134	206	coefficient	132	207	deposition	131
208	procedure	131	209	internal	129	210	evaluate	128
211	previously	128	212	specifically	128	213	variable	128
214	domain	127	215	feature	127	216	interaction	127
217	amplitude	125	218	cycle	124	219	feasibility	123
220	friction	123	221	setup	123	222	via	123
223	implemented	122	224	cell	121	225	indicates	121
226	predictions	121	227	resolution	121	228	whereas	121
229	inverse	120	230	plastic	120	231	static	120

232	tissue	120	233	varying	120	234	evaluation	119
235	generation	119	236	utilized	119	237	removed	117
238	global	116	239	offset	116	240	additionally	115
241	decomposition	115	242	external	115	243	precision	115
244	tangential	115	245	ensure	114	246	estimate	114
247	modified	114	248	rigid	113	249	challenges	112
250	define	112	251	equivalent	112	252	manual	112
253	prior	112	254	structural	112	255	focus	111
256	illustrated	111	257	experimentally	109	258	novel	109
259	optical	109	260	controllers	108	261	mathematical	108
262	contrast	107	263	indicate	107	264	theoretical	107
265	vibration	107	266	images	106	267	investigate	106
268	theory	106	269	vectors	106	270	traditional	105
271	similarly	104	272	issues	103	273	researchers	103
274	finite	102	275	variations	101	276	adaptive	100
277	consistent	100	278	evolution	100	279	nonlinear	100
280	stable	100	281	minimize	99	282	validation	99
283	consists	98	284	porosity	97	285	analytical	96
286	areas	96	287	occurs	96	288	series	96
289	simulated	96	290	expansion	95	291	regions	95
292	accelerometer	94	293	affected	94	294	identification	94
295	integration	93	296	optimized	93	297	tasks	93
298	enable	92	299	flexibility	91	300	assumption	89
301	computed	89	302	differential	89	303	embedded	89
304	etching	89	305	nominal	89	306	modules	88
307	sufficient	88	308	uniform	88	309	versus	88
310	analyzed	86	311	digital	86	312	rotational	86
313	sections	86	314	vertical	86	315	alternative	85
316	appropriate	85	317	authors	85	318	real-time	85
319	simulations	85	320	transformation	85	321	accurately	84
322	assuming	84	323	environments	84	324	established	84
325	focused	84	326	implementation	83	327	trend	83
328	carrier	82	329	normalized	82	330	occur	82
331	overlap	82	332	empirical	81	333	identical	81
334	modulus	81	335	role	81	336	bandwidth	80
337	framework	80	338	peak	80	339	threshold	80
340	electron	79	341	located	78	342	summarized	78
343	capable	77	344	detection	77	345	methodology	77
346	analog	76	347	polymer	76	348	conclusions	75
349	construction	75	350	dimensional	75	351	maintain	75
352	plotted	75	353	primary	75	354	distributed	74
355	microscope	74	356	simultaneously	74	357	constructed	73
358	indicating	73	359	induced	73	360	specified	73
361	capability	72	362	plot	72	363	responses	72
364	transient	72	365	creating	71	366	etched	71
367	issue	71	368	conclusion	70	369	instance	70
370	velocities	70	371	aluminum	69	372	architecture	69
373	diagram	69	374	network	69	375	morphology	68

376	vary	68	377	capabilities	67	378	estimation	66
379	extracted	65	380	robust	65	381	validate	64
382	conductivity	63	383	attached	62	384	chemical	62
385	compliance	62	386	pitch	62	387	remove	62
388	compensation	60	389	hydraulic	60	390	micro	60
391	potentially	60	392	validated	60	393	monitoring	59
394	cells	58	395	appendix	56	396	carbide	53
397	mounted	51	398	microscopy	44			

Appendix 2

Sublist 1 in AWL (in alphabetical order)

analyse	define	indicate	proceed
approach	derive	individual	process
area	distribute	interpret	require
assess	economy	involve	research
assume	environment	issue	respond
authority	establish	labour	role
available	estimate	legal	section
benefit	evident	legislate	sector
concept	export	major	significant
consist	factor	method	similar
constitute	finance	occur	source
context	formula	percent	specific
contract	function	period	structure
create	identify	policy	theory
data	income	principle	vary

(Source: Coxhead, 2000, p.232-235)

Acknowledgement

1. This study was kindly supported by a grant of 2020 Research Funds from the Educational Department of Liaoning Province, China (Project No. LG202015). The comments of the three anonymous reviewers are highly appreciated.
2. Part of the results in the study is also represented in the author’s article of 2022 in *Corpus Linguistics*, *9*(1),150-160.

References

Brezina, V. & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*,1-22.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-238.

Coxhead, A. & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, *16*, 129-147.

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

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, *43*, 49-61.

Ferrell, P. (1990). Vocabulary in ESP: A lexical analysis of the English of Electronics and a study of semi-technical vocabulary. *CLCS Occasional Paper No. 25*. Trinity College. (ERIC Document Reproduction Service No. ED332551).

Gardner, D. & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*, 305-327.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, *33*, 54-65.

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

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*, 4-21.

Hyland, K. & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*, 235-254.

Jiang, S. J. (2010). Creating petroleum academic word list. *Journal of Southwest Petroleum University*, *3*(6), 27-30.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51.

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, *39*(1), 1-11.

Liu, D. L., & Lei, L. (2020). A review of the study on academic word lists. *Foreign Language Education*, *41*(2), 34-50.

Martinez, I. A., Beck, S. C., & Panza, C. B. (2009), Academic vocabulary in agriculture research articles. *English for Specific Purposes*, *28*(3), 183-198.

McEnery, A., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh University Press.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*, 235-256.

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 1-12.

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.

Schmitt, N., Cobb, T., Horst, M. & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, *48*(2), 212-226.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Valipouri, L. & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, *12*, 248-263.

Veenstra, J. & Sato, Y. (2018). Creating an institution-specific science and engineering academic word list for university students. *The Journal of Asia TEFL*, *15*, 148-166.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*, 170-182.

Webb, A. S., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*, 113-126.

West, M. (1953). *A general service list of English words*. Longman, Green.

Wu, J., & Wang, T. S. (2007). Researching the applicability of Coxhead's academic word list. *Foreign Language Teaching Abroad*, *27*(2), 28-33.

Zhao, Z. G. (2015). Development of ESP academic word lists: A case study of maritime English. *Journal of Chongqing Jiaotong University*, *15*(6), 140-144.

Le Chang received his PhD in Language Teaching and Learning from the University of Auckland, New Zealand in 2013. He is presently head of the Applied Linguistics Research Centre and a professor of English at Shenyang Ligong University, where he teaches EAP to mechanical engineering graduates. He is also an MA supervisor (Foreign Linguistics & Applied Linguistics) at Bohai University. His research interests include EFL vocabulary acquisition, EFL listening teaching, and corpus-based EAP teaching.