The Development of an Academic Collocation List for Undergraduate Mechanical Engineering Students

BOONYAKORN SIENGSANOH
Chulalongkorn Language Institute, Bangkok, Thailand
Author email: paiboonyakorn@gmail.com

Article information

Article history:
Received: Mar 8, 2021
Accepted: Apr 2, 2021
Available online: Apr 9, 2021

Keywords:
Technical vocabulary
Collocation
Lexical collocation
Mechanical Engineering corpus
ESP
Vocabulary list

Abstract

Acquiring collocations is widely considered to be essential in language learning. The question arises, however, as to what collocations merit students' and teachers' attention in ESP classes at a tertiary level. Thus, this corpus-based study aims to systematically develop an academic collocation list for undergraduate mechanical engineering students who learn English as a foreign language. To identify such a collocation list, the current study relied on a corpus-based approach and an expert-judged approach. The Sample Corpus of Mechanical Engineering containing 2.1 million words was compiled from required and supplementary textbook chapters, reading texts and research articles as specified in the course syllabi for the undergraduate program in Mechanical Engineering at a public university in Thailand. The development of the list involved five stages: 1) compilation of materials 2) creating a specialized corpus 3) extracting high-frequency node words and identifying lexical collocations 4) expert judgement; and 5) ordering the entries. These steps are in line with those proposed by Ackermann and Chen (2013) with some modifications. It is expected that this corpus-informed collocation list consisting of 282 entries will be highly useful for students majoring in Mechanical Engineering as well as ESP teachers and material developers. The complete list of collocations is provided in the appendix.

INTRODUCTION

According to Gardner and Davies (2013), academic vocabulary knowledge is essential for academic reading ability and also for learners' academic success. Therefore, various lists of academic words were created: West's (1953) General Service List (GSL), Xue and Nation's (1984) University Word List (UWL), Coxhead's (2000) Academic Word List (AWL), Gardner and Davies' (2013) Academic Vocabulary List (AVL), Brezina and Gablasova's (2013) New General Service List, to name a few. These lists have served as a useful source for vocabulary learning as they can be used directly by learners or help teachers and material developers when they produce in-house materials or design English for Academic Purposes (EAP) or English for Specific Purposes (ESP) courses. Coxhead (2000, p. 214) affirmed that "an academic word list should play a crucial role in setting vocabulary goals for language courses, guiding learners in their independent study, informing course and material designers in selecting texts and developing learning activities." However, some educators (e.g., Chung & Nation, 2004) argue that a more specialized word list, also known as a technical word list, is also necessary. To highlight the importance of a discipline-specific vocabulary list, Durrant (2009) argued that the vocabulary needs of learners in an academic discipline should be characteristically different from those in other disciplines. Thus, teachers have to deal with these student needs separately.

In order to investigate academic vocabulary that merits students' attention, whether it be general or discipline-specific academic words, many studies have relied on a corpus-based approach. For instance, Coxhead's (2000) Academic Word List (AWL), which consists of 570 word families, was obtained from a 3.5-million-word corpus of written academic text. Gardner and Davies (2013) explored the 120-million-word academic subcorpus of Corpus of Contemporary American English (COCA) to identify 500 words (or lemmas, to be precise) for the Academic Vocabulary List (AVL). In addition, in an attempt to identify specialized words in natural science disciplines, It-ngarm and Phoocharoensil (2019) created a 5.5-million-word corpus called the Science Academic Journal (SAJ) Corpus and extracted 432 word families that were frequently found. The following section will discuss reasons why corpora have been widely used among researchers who developed EAP and ESP word lists.

Corpora as a tool for creating word or collocation lists

It has been well-documented that corpora can be used to improve the way ESP teaching is approached (e.g., O'Keeffe & McCarthy, 2010; Mudraya, 2006; Coxhead, 2000). Toriida (2016, p.89) pointed out that "a corpus-based approach is a form of evidence-based language pedagogy that provides teachers with information to guide decisions regarding vocabulary teaching, learning, and testing." To illustrate another benefit, McEnery and Wilson (2001) maintained that corpora provided domain-specific materials for language learning, which meet the needs of ESP students. The use of corpora also allows researchers to compare the frequency of a set of vocabulary or word combinations in a particular field of study with that in other 'more general' reference corpora. That is why several studies have focused on developing corpus-based discipline-specific vocabulary lists. However, this line of research mostly focused on creating lists of individual words and most of them did not create the lists based on a curriculum of a discipline. The present study therefore aims to bridge this gap through focusing on technical collocations in Mechanical Engineering.

The aims of this study are:

1) to create a curriculum-based representative corpus of Mechanical Engineering which presents frequent discipline-specific collocations.
2) to develop a corpus-based Mechanical Engineering collocation list which is highly useful for undergraduate Mechanical Engineering students and ESP teachers.

Significance of learning and acquiring collocations

There are many reasons why the significance of collocations, also called word combinations, prefabricated chunks (or prefabs), phraseological units, multiword units, or formulaic sequences, has been well-acknowledged. First, according to Chon & Shin (2013), formulaic sequences make up a large proportion of natives' lexical knowledge. It is believed that collocational knowledge is defining markers of near-nativeness. So, in the L2 learning contexts, it is very important for learners to be able to comprehend and retrieve lexical items as in prefabricated chunks. This seems to be consistent with what McCarthy and O'Dell (2005) suggested in the introduction of their book: increasing collocational knowledge can help learners speak and write English more naturally and precisely. Those who would like to specialize in a particular academic field need to sharpen their skills in selecting an appropriate word that fits the context, and this entails an ability to use collocations properly. If learners use inappropriate collocations, they may sound unnatural or, in some cases, even incomprehensible to others. Despite this fact, Ackermann and Chen (2013) note that producing collocations is a huge challenge for learners partly because they usually rely on a limited number of collocations. Also, some learners tend to overuse some sequences they are familiar with and underuse some less frequent but strongly associated collocations (Chon & Shin, 2013). In terms of receptive skills, some learners seem to be unable to distinguish a pair of near-synonyms which, according to Biber and Conrad (1999), will be easier to recognize if they know the collocations of those words. Because of these problems, it would be pedagogically valuable to have a collocation list which could help learners expand their vocabulary knowledge and also assist ESP teachers in making well-informed decisions about which word combinations merit their attention during their class.

METHODOLOGY

Research procedure

In an attempt to develop a corpus-based collocation list, the methods proposed by Ackermann and Chen (2013) were used with some adjustments. This involved five steps: 1) compiling course syllabi and preparing materials for corpus 2) creating a specialized corpus 3) extracting high-frequency node words and identifying lexical collocations; and 4) expert judgement; and 5) ordering the entries. In this section, the construction of a specialized corpus, The Sample Corpus of Mechanical Engineering, is described in detail together with the collocation selection criteria. Then the list's implications for teaching and for course material development will be discussed. Also, future research needs will be outlined.

Step 1: Compiling course syllabi and preparing materials for corpus

Twenty-three course syllabi of all subjects for the undergraduate program of Mechanical Engineering at a public university in Thailand in Academic Year 2019 were compiled and categorized into three sub-disciplines: 1) Fluid Mechanics and Thermodynamics, 2) Dynamics and Control, and 3) Solid Mechanics. To ensure the accuracy of classification, a full-time lecturer at the Department of Mechanical Engineering at the university assisted in classifying each subject into the three sub-disciplines. Since the study particularly focused on Mechanical Engineering, all general or basic subjects for engineering students, including non-Mechanical Engineering students, were excluded at this stage. These general subjects are, for instance, Engineering Drawing, Engineering Measurement, and Modern Computer-Based Manufacturing

System. The required and supplementary textbooks, articles, and reading texts as specified in the course syllabi were subsequently compiled and saved in PDF formats so that they were compatible with LancsBox, which was a concordancer employed in the next step. To do this, a great deal of attention must be paid when the book chapters were selected. It is worth noting that all the compiled materials were a complete chapter or section in the books, and not excerpts. Furthermore, to exclude the materials irrelevant to Mechanical Engineering i.e. to ensure the corpus consists of representative texts, some sections of the books such as prefaces, contents, acknowledgements, author's biography, indexes, references or bibliographies, and appendixes were removed manually during this stage.

Step 2: Creating a specialized corpus

In order to create a corpus with a balanced range of topics, the materials selected were under the three sub-disciplines of Mechanical Engineering with equal weight. This is the technique used by various previous studies on vocabulary lists (e.g., It-ngarm & Phoocharoensil, 2019). The number of tokens per sub-discipline and the total number of tokens of The Sample Corpus of Mechanical Engineering together with its percentage are provided in Table 1.

Table 1
Number of tokens per sub-discipline

Sub-disciplines of Mechanical Engineering	Number of tokens (percentage)
1. Fluid mechanics and thermodynamics	705,167 (33.34%)
2. Dynamics and control	705,890 (33.38%)
3. Solid mechanics	703,907 (33.28%)
Total Number of Tokens	2,114,964

Step 3: Extracting high-frequency node words and identifying lexical collocations

3.1 Concordancer

LancsBox v.5.1.2, developed by Brezina, Weill-Tessier and McEnery (2020), was used because it is equipped with features for counting tokens, types, and lemmas, calculating frequency of words, and extracting collocations, which serve the purpose of the current study.

3.2 Identifying high-frequency node words

Since the collocation list was created based on the assumption that frequency should be an indicator of whether a word would be useful for learners and teachers (Ackermann & Chen, 2013; Coxhead, 2000; Mudraya, 2006; Valipouri & Nassaji, 2013), the first criteria for identifying node words was frequency. The identification started with using the Words function in LancsBox v.5.1.2 to extract 3,500 most frequent words. Then, the 3,500 words underwent a manual qualitative review to exclude the following types of words:

a. Numbers, abbreviations, acronyms (e.g., EOUT, ASME, ASTM, MOSFET), and non-words such as one-letter or two-letter words which can be a part of math formulas, measurement units (e.g., psia, kpa, amp), or variables in equations.

b. Proper names (e.g., American). However, it seems pretty common in the Mechanical Engineering field that some principles or concepts were named after influential engineers, physicists, or scientists, so it was decided that the proper names which are a part of well-established principles or concepts (and which occured with high frequency) were kept.

c. Function words such as preposition, determiners, conjunctions, pronouns, modals, and question words. Thus, only content words were included.

During this stage, the words also went through manual lemmatization to group together the inflected forms of a word in the same word family. The form which occurred most frequently was selected as a node word for the next step. To illustrate how the manual lemmatization was performed, consider the following frequency outputs from LancsBox v.5.1.2.

Table 2
Sample frequency output of the 'refrigerant' word family

Word form	Absolute frequency of a word form in the corpus
refrigerant	**203**
refrigerated	54
refrigeration	171
refrigerator	186
refrigerators	46

In this case, it is obvious that, compared to the other word forms, 'refrigerant' is most frequently found in the corpus. Thus, 'refrigerant' was selected as a node word.

Table 3
Sample frequency output of the 'rotate' word family

Word form	Absolute frequency of a word form in the corpus
rotate	205
rotated	134
rotates	186
rotating	380
rotation	**548**
rotational	177
rotations	118

Considering the frequency output above, ‘rotation’ is much more frequent than the other word forms, so ‘rotation’ was selected as a node word.

Following the lemmatization, General Service List (West, 1953) and Academic Word List words (Coxhead, 2000) were excluded from the list using the Microsoft Excel software program. Another criterion set for the node word selection was the absolute frequency of occurrence of > or = 60. This criterion was applied so that the final list would not be too long and readily applicable in the classroom. In addition, this is to follow Coxhead’s (2000) AWL word selection criteria. In order for a word to be included in the AWL list, it must occur over 100 times in Academic Corpus containing 3.5 million tokens. Because The Sample Corpus of Mechanical Engineering contains 2.1 million tokens (2,114,964, to be precise), a word which occurred at least 60 times in the corpus was selected as a node word. In this step, a great number of high-frequency words were excluded and the resulting list was reduced to 379 node words.

3.3 Identifying lexical collocations

According to Hunston (2002, p. 68), “It [collocation] can be considered as the tendency of two words to co-occur, or as the tendency of one word to attract another.” Also, Hunston (2002) highlighted the importance of using a corpus to investigate collocations by arguing that “collocation may be observed informally in any instance of language, but it is more reliable to measure it statistically, and for this a corpus is essential” (p. 68). The node words obtained in Step 3.2 were used to extract the collocates. Below are the criteria for determining collocates for the study.

a. The word must be in the 3-word span on the right or left of the node words. This is because the range is not too far nor too close for each pair to co-occur.

b. The collocate must be a complete word, not a number, an abbreviation or an acronym.

c. Lexical collocations, not grammatical ones, were selected.

d. To take the association strength (or collocational strength) of each pair into account, the Mutual Information (MI) score of the pair must be > or = 3, which is the recommended score by Hunston (2002). Ackermann & Chen (2013) adopted this criterion as well. An MI-score “compares the actual co-occurrence of the two items with their expected co-occurrence if the words in the corpus used were to occur in totally random order. In other words, the MI-score measures the amount of non-randomness present when two words co-occur” (p. 71).

e. The threshold, or the minimum frequency of the collocation (i.e. the occurrence of each pair) is set to be > or = 5.

In an effort to extract the collocate of each node word which met all the criteria above, the GraphColl tool in LancsBox v.5.1.2 was used. After setting the word span, the Mutual Information (MI) score, and the threshold, the GraphColl tool identified the collocates of each node word and ranked them in the order of decreasing MI scores as illustrated in the table below.

Table 4
Collocate output of 'amplitude'

Position	Collocate	MI score	Frequency of collocation
R	linearity	11.71	16
L	modal	10.09	22
R	scaling	9.99	5
R	decay	9.90	5
R	sinusoidal	8.66	5

As seen in the table, the collocate with the highest MI score were ranked first, followed by that with the second highest MI score and so on. The collocate which occurred most frequently among the top five collocates were included in the list. Therefore, the selected collocate of 'amplitude' was 'modal,' the position of which was on the left of the node word. In case a pair co-occurs as frequently as another, a pair with the higher MI score was chosen.

Any word combinations which did not meet these criteria were eliminated. There were, however, a few cases in which the node words were excluded, but not because they did not meet the set criteria. Rather, it was because of other reasons which are worth mentioning here as it would be beneficial for future research on collocation lists. First, as shown by their statistic information, a few node words (e.g., coil and array) usually stand alone i.e. they rarely appear with any collocates and/or are often surrounded with function words. These kinds of words were excluded during this process. Second, pairs of words which were obviously irrelevant to Mechanical Engineering were put aside. They were, for instance, *photo courtesy, mentioned earlier* and *downloaded (from a) website*. Supposedly, these pairs of words were more related to giving credit to a photographic content provider, or to referring back to a previous section of the book, but not to Mechanical Engineering, which is the main focus of the current study. Thus, they were not included in the list. Finally, it is also noteworthy that when two node words, which met all the criteria discussed earlier (high frequency and high MI score) happened to collocate with each other, it was decided that the pair of words was presented in the list once. Such word combinations were, for example, *combustion chamber, Newton's law, free-body diagram, oxygen (and/or) nitrogen, radius (of) gyration, sleeve bolt, static friction,* and *Cartesian vector*. This was the reason why the resulting list was reduced to 333 pairs of words.

Step 4: Expert judgement

According to Chung and Nation (2004, p. 252), "technical vocabulary is part of a system of subject knowledge. It could thus be identified by referring to specialists who have a good knowledge of the subject area." In agreement with this statement, a panel of five experts in Mechanical Engineering with at least six years of teaching experience (average = 15.6 years) at the Department of Mechanical Engineering at a public university were requested to judge whether each pair of words in the potential list should be included in the final list or not. The purpose of the expert review was to find out if the entries, which satisfied the aforementioned quantitative criteria were really worth teaching from a pedagogical perspective (Ackermann

& Chen, 2013). The panel experts were informed of the objective and scope of the study, and were provided with statistical information about each entry i.e. the absolute frequency of each node word, the MI-score of each pair, and the absolute frequency of each pair in the 2.1-million-word corpus. They were then requested to rate the possible collocations, using the four-level scales. Below are the labels for the four-point Likert scale which were used by Ackermann & Chen (2013) and were partially based on the guidelines about using a rating scale for technical word identification suggested by Chung and Nation (2004).

1 = definitely exclude
2 = not sure, but tendency to exclude
3 = not sure, but tendency to include
4 = definitely include

The entries that were rated 1 or 2 by two of the five experts were removed from the list. During this process, the experts suggested that additional contexts should be added to some pairs of words so as to make the meanings much clearer. The researcher found the suggestion highly constructive and justified because the contexts would make the list even more readily applicable in classroom as well. Thus, coordinating conjunctions (i.e. and/or) or prepositions (e.g., of a, on) were added and put in parenthesis in order to prevent confusion when the list is used. Moreover, proper names were capitalized during this step.

Step 5: Ordering the entries

There are many ways to organize or order the entries in the list. While some researchers (Coxhead, 2000; Valipouri & Nassaji, 2013) believed that the organization should be based on frequency, Thornbury (2002) argued for using high learnability (i.e. easy to learn) and teachability (i.e. easy to teach) to organize the list. The rationale behind this idea is teaching efficiency in classroom. In addition, Watson Todd (2017) proposed that opacity of words should be another consideration when sequencing the list. To help teachers make an informed decision on what words merit more attention in a classroom, it was recommended that opaque words be put at the top of the list, whereas more transparent words should be put at the bottom. He convincingly argued that it would be a good idea to devote limited classroom time to words whose meanings were opaque or difficult for students to understand on their own.

In ordering the entries for the present study, Coxhead's frequency criterion was adopted and the collocation list (see the Appendix) was presented in decreasing order of frequency of the node words to facilitate teachers. Teachers can probably start from teaching the collocations which occurred more frequently first.

RESULTS AND DISCUSSION

After five stages of corpus analysis and manual qualitative refinement described earlier, the academic collocation list for undergraduate Mechanical Engineering students was developed. The final list consisted of 282 entries. Further investigation into the absolute frequencies of the node words showed that the highest frequency was 2,831 and the lowest was 60 times in The Sample Corpus of Mechanical Engineering, which comprised 2.1 million tokens. The absolute frequencies of the pairs of words ranged from 677 to 5 times. Since the high-frequency node words were ranked first in the list, it is recommended that undergraduate Mechanical Engineering students and ESP teachers at the tertiary level focus more on these collocations than those ranked lower. Also, since the list contains a pair of words, as opposed to single words, it is hoped that when the students use this list to improve their vocabulary knowledge, they would also be constantly reminded of the importance of both structural and lexical meanings, which are two elements of the English language we need to truly understand when producing or comprehending the language.

Another possible application of the list is for ESP material developers. It would be highly beneficial for ESP learners if high-frequency collocations are incorporated in teaching and learning materials, or alternatively, in classroom activities. Because the list was also proved to be pedagogically valuable from Mechanical Engineering experts' point of view, ESP material developers and teachers can rest assured that familiarizing students with these pairs of words would not be a wasteful use of time and resources. Rather, with this study, such instruction would be well-grounded.

Furthermore, this study has enhanced our understanding of how to develop a specific-discipline collocation list. Several observations will be discussed here. First of all, several words in General Service List developed by West in 1953 (e.g., *flow, absolute,* and *contract*) were found to be polysemous and some of their meanings could be more useful for Mechanical Engineering students than others. That is why the idea of focusing on opaque words proposed by Watson Todd (2017) was reasonable and those words merit students' and teachers' attention as well.

Second, during the process of excluding words from General Service List and Academic Word List, it was found that certain words may seem to come from the same word family as those in two lists. However, upon closer examination, they actually have very specific and technical meanings, and therefore deserved to be included in the list of selected node words. These interesting words include *actuator, airflow, bandwidth, centerline, circuit, stainless,* and *static.* At first, these words look like GSL words which are *act, air, band, center, circle, stain,* and *state* respectively. Also, at first glance words like *analog, automobile, compensator, concentric,* and *projectile* might look like AWL words including *analogy, automatic, compensate, concentrate,* and *project,* respectively.

Third, it should be noted that some word forms are more common in Mechanical Engineering than the others in the same word family. This linguistic evidence can be considered characteristic of technical English. The tables below can illustrate this point.

Table 5
Absolute frequencies of the 'react' word family

Word	Absolute frequency
react	41
reactants	129
reacting	46
reaction	500
reactions	358

Table 6
Absolute frequencies of the 'result' word family

Word	Absolute frequency
result	1094
resultant	668
resultants	79
resulted	40
resulting	412
results	1012

As seen in the tables above, *reaction* and *reactions* are found to occur much more frequently in the Sample Corpus of Mechanical Engineering than *reactants, reacting,* and *react.* As for the 'result' word family, *result, results,* and *resultants* appear a lot more frequently than *resulting, resultants,* and *resulted*. These findings could be useful for both ESP teachers and Mechanical Engineering students. Such findings can also serve as base for future research. In addition, while examining the list of word forms, it was brought to the researcher's attention that oftentimes, nouns are most frequently found compared to other word forms. Take Tables 3, 5, and 6 above as examples. Nevertheless, further investigations are required to validate this point.

Limitation and future research

One limitation of the study lies in the fact that the materials compiled in the Sample Corpus of Mechanical Engineering were limited to textbook chapters, reading text, and research articles specified in the course syllabi only. It is recommended that future research explore other kinds of learning materials such as slides or lectures in order to gain more insight into pedagogically valuable ESP words or collocations.

CONCLUSION

This corpus-based collocation list can be applied in settings where the instruction focuses on improving ESP knowledge of undergraduate students. It equips teachers with vocabulary knowledge necessary for them to "speak the same language" as their ESP students. Based on corpus analysis and experts' evaluation, this list can help teachers and students decide which sets of vocabulary or collocations should be prioritized and incorporated in learning materials. It is hoped that this list would be like a compass which helps navigate students through their journey of mastering ESP English as well as gaining expertise in Mechanical Engineering, and ultimately lead them to an academic success.

THE AUTHOR

Boonyakorn Siengsanoh has been working as an English lecturer at Chulalongkorn University Language Institute for seven years. She completed her M.A. in English Linguistics at the Faculty of Arts, Chulalongkorn University. Her areas of research interests include vocabulary teaching strategies, L2 writing, collocations, and corpus linguistics.
paiboonyakorn@gmail.com

REFERENCES

Ackermann, K., & Chen, Y. H. (2013). Developing the academic collocation list (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora* (pp. 181-190). Amsterdam, The Netherlands: Rodopi.

Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*(1), 1-22.

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: http://corpora.lancs.ac.uk/lancsbox.

Chon, Y. V., & Shin, D. (2013). A corpus-driven analysis of spoken and written academic collocations. *Multimedia-Assisted Language Learning, 16*(3), 11-38.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System, 32*(2), 251–263. https://doi.org/10.1016/j.system.2003.11.008

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

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for Academic Purposes. *English for Specific Purposes, 28*(3): 157-169.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, England: Cambridge University Press.

It-ngam, T., & Phoocharoensil, S. (2019). The development of science academic word list. *Indonesian Journal of Applied Linguistics, 8*(3), 657–667. https://doi.org/10.17509/ijal.v8i3.15269

Lei, L., & Liu, D. (2018). The academic English collocation list. *International Journal of Corpus Linguistics, 23*(2), 216–243. https://doi.org/10.1075/ijcl.16135.lei

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Boston, MA: Thomson.

McCarthy, M., & O'Dell, F. (2005). English collocations in use. Cambridge, England: Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed). Edinburgh, Scotland: Edinburgh University Press.

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

O'Keeffe, A., & McCarthy, M. (Eds.) (2012). *The Routledge Handbook of Corpus Linguistics*. New York, NY: Routledge.

Thornbury, S. (2002). *How to teach vocabulary*. London, England: Pearson.

Toriida, M.-C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list, *TESOL Canada Journal, 34*(11), 87-105. https://doi.org/10.18806/tesl.v34i1.1257

Veenstra, J., & Sato, Y. (2018). Creating an institution-specific science and engineering academic word list for university students. *Journal of Asia TEFL, 15*(1), 148–166. https://doi.org/10.18823/asiatefl.2018.15.1.10.148

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. https://doi.org/10.1016/j.jeap.2013.07.001

Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*(1), 31- 39. https://doi.org/10.1016/j.esp.2016.08.003

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001

West, M. (1953). *A general service list of English words*. London, England: Longman, Green & Co.

Wudthayagorn, J., Pugsee, P., Janpugdee, P., & Limgomolvilas, S. (2020). A study of the information and communication technology word list. *Pasaa Paritat Journal, 35*, 145-167. Retrieved from https://www.culi.chula.ac.th/publicationsonline/home_p2.php.

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215-219.

Appendix

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
1	angular	velocity		2831
2	exact	solution		2134
3		plane	symmetry	1624
4	free-body	diagram		1597
5		thermal	conductivity	1596
6	working	fluid		1546
7	gravitational	acceleration		1486
8	octahedral	shear		1440
9		angular	momentum	1406
10	first-stage	turbine		1396
11	maintain	equilibrium		1340
12	assembled	matrix		1286
13	moments (of)	inertia		1277
14		magnitude	(and/or) direction	1222
15		fuel	consumption	1042
16		horizontal	(and/or) vertical	1023
17		strain	gauge	987
18		vertical	plane	985
19	composite	cylinder		978
20	static	friction		978
21		linear	momentum	918
22		circuit	board	904
23	saturated	vapor		893
24	convection	coefficient		873

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
25		relative	humidity	851
26		voltage	regulator	839
27	tension (and/or)	compression		790
28		feedback	system	790
29		particle	moves	787
30		entropy	production	777
31	impeller	diameter		769
32	pin-ended	column		752
33		kinetic	energy	699
34	Bode	plot		673
35	root	locus		629
36		saturated	liquid	601
37	relief	valve		579
38		fatigue	growth	578
39	translation (and/or)	rotation		548
40		tank	contains	532
41	turbine	exit		529
42		mesh	generation	517
43		condenser	(and/or) evaporator	495
44		disk	drive	494
45		axial	rigidity	483
46		density	(and) viscosity	477
47		isentropic	efficiency	460
48	rotating	shaft		447
49	roof	truss		425
50		enthalpy	formation	415

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
51	nodal	displacements		410
52		plastic	zone	407
53	transient	conduction		400
54	viscous	damping		399
55		poles	(and/or) zeros	392
56		closed-loop	poles	390
57	reciprocating	engines		389
58		torque	applied	376
59		loop	transfer (function)	371
60	directed	perpendicular		364
61		nonlinear	simulation	351
62		Cartesian	vector	349
63	lateral	deflection		347
64		carbon	monoxide	346
65		software	packages	343
66	combustion	chamber		332
67		sketch	locus	330
68	center (of)	gravity		323
69		adiabatic	flame (temperature)	319
70	test	specimen		319
71	modal	amplitude		315
72		tensile	(or) compressive	311
73	low-pass	filter		308
74		cross-sectional	area	305
75		hydraulic	cylinder	301
76		exhaust	gases	288

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
77		three-dimensional	mesh	287
78		mathematical	models	276
79	planar	kinematics		275
80		feedwater	heater	271
81	forced	vibration		266
82		Carnot	cycle	265
83		impulse	momentum	264
84		modulus	elasticity	263
85	transer (to)	surroundings		263
86	relative	humidity		259
87	plane (or axis)	symmetry		259
88		atmospheric	pressure	258
89		piston	moves	258
90	specimen	geometry		250
91	propel (of an)	aircraft		242
92		magnetic	flux	239
93		torsion	box	239
94		centroid	area	237
95		exerted	force	237
96	aluminum	alloy		233
97		Rankine	cycle	233
98		gear	rack	228
99		analog	signal	225
100	user	interface		225
101		nozzle	exit	222
102	local	buckling		218

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
103		actuator	(and/or) sensor	214
104		elongation	bar	210
105	sleeve	bolt		207
106		conservation	principle	207
107		fracture	toughness	207
108		refrigerant	enters	203
109	radius (of)	gyration		200
110	pushbutton	switch		199
111	percent	overshoot		198
112		polynomial	degree	195
113	explicit	finite-difference		194
114	well	insulated		192
115		longitudinal	strains	189
116	inlet (and/or)	outlet		188
117	belt (and/or)	pulley		185
118	Newton's	laws		184
119		patch	test	184
120		corrosion	(and/or) erosion	182
121		tangent	path	181
122		lateral	deflection	180
123		viscous	coefficient	177
124		homogeneous	solid	175
125		diesel	engines	174
126		interior	node	173
127	compound	pendulum		173
128	bipolar	transistor		171

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
129	oxygen (and/or)	nitrogen		170
130		microcontroller	programming	166
131		fin	('s) profile	161
132	analog-to-digital	converter		160
133		cord	tension	159
134		flange	face	159
135		Laplace	transform	159
136		incompressible	substances	158
137	vertically	upward		156
138		fraction	expansion	154
139		Lagrange	multipliers	152
140		robust	control	152
141		automobile	engines	151
142		orthogonal	directions	151
143		slender	rod	149
144	plastic	deformation		146
145	(light-)emitting	diode		145
146	gas	furnace		145
147	threaded	welding		143
148		algebraic	signs	142
149		Brayton	cycle	142
150		gage	pressures	142
151		transverse	section	140
152		mechatronic	systems	139
153		capacitor	(and/or) inductor	137
154	load	platform		137

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
155	(exposed to)	ambient	air	136
156	carbon	dioxide		136
157		elevation	drop	136
158		modal	vectors	135
159		one-dimensional	conduction	135
160		threshold	comparator	132
161		multiple	degrees of freedom	128
162		thrust	bearing	126
163		gas-turbine	engines	125
164	solved	simultaneously		125
165	input	impedance		124
166		ramp	input	124
167		stainless	steel	124
168		diagonal	matrix	119
169	bipolar	junction		119
170		moisture	content	118
171		inverse	transform	117
172		open-loop	gain	117
173	smart	grid		116
174		threaded	joint	114
175		isothermal	expansion	113
176		superheated	vapor	113
177		upstream	flow	112
178	model (and/or)	prototype		111
179	wind	tunnel		111
180		residual	stresses	110

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
181		binary	(and/or) decimal	109
182		infinitesimal	element	109
183		supercritical	flow	109
184		jet	engine	107
185		lab	exercise	107
186		verify	results	107
187		cast	iron	106
188	quadratic (and/or)	cubic		106
189		metals	(and) alloys	106
190	energy	dissipation		105
191		electronic	device	105
192	thick	slab		105
193		satellite	attitude	104
194		prescribed	conditions	102
195		parametric	plane	101
196	simple	harmonic	(motion)	100
197		multidegree-of-freedom	systems	99
198		ideal-gas	behavior	97
199		first-order	system	96
200		prismatic	bar	96
201	liquid	propane		96
202		lumped	capacitance (method)	95
203		subcritical	flow	95
204		thin-walled	members	95
205	source (and/or)	drain		94

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
206		helium	gas	93
207	statically	indeterminate		92
208	serial	port		92
209	surface	profile		92
210	rear	axle		91
211	manufacturing	semiconductor		89
212	frequency	bandwidth		88
213	time	increment		88
214	design	compensator		87
215		interpolation	functions	87
216		nonzero	terms	87
217		projectile	motion	86
218		cantilever	beam	85
219		curvilinear	motion	85
220	numerator (and/or)	denominator		85
221		stoichiometric	amount	85
222	absolute	vacuum		85
223		constant-pressure	heat addition	83
224		sinusoidal	response	83
225	silicon	chip		82
226	smooth	slot		82
227		superposition	principle	82
228		high-temperature	reservoir	81
229		eigenvalue	problem	80
230		notch	filter	80
231		compliance	matrix	79

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
232		Reynolds	number	79
233		dashed	line	78
234		electromechanical	dynamics	78
235	microwave	oven		78
236		potentiometer	sensors	78
237		reciprocating	compressors	78
238	crack	propagation		77
239		constant-volume	addition	76
240		isotropic	material	76
241		open-channel	flow	75
242		spool	rolls	75
243		gasket	material	74
244		liquids	(and/or) solids	74
245		octahedral	stress	73
246	compression	stroke		73
247		tuning	method	73
248		ductile	iron	72
249	pressure	gradient		72
250		Froude	number	71
251	sulfuric	acid		69
252		air-standard	cycle	69
253	frequency	oscillation		69
254		pivot	point	69
255		solar	panels	69
256	chip (on)	substrate		69
257		solder	joint	69
258		air-conditioning	system	68
259		dew-point	temperature	68

No.	Pre-Collocate	Node words	Post-Collocate	Absolute frequency of node words
260		inelastic	columns	67
261		rectilinear	kinematics	67
262	passenger	compartment		66
263		parallel-axis	theorem	66
264		pitch	rate	66
265		compact	heat (exchanger)	65
266		intake	(and) exhaust	65
267		traction	motor	65
268		aerodynamic	drag	64
269	(US)	customary	units	64
270		laser	beam	64
271	inertia	ellipsoid		63
272	circular	orbit		63
273	spark	plug		62
274		semi-infinite	solid	62
275		silicon	diode	62
276		crank	rotates	61
277	point (of)	intersection		61
278	height (of)	mercury		61
279		parabolic	trajectory	61
280		axisymmetric	element	60
281		low-temperature	reservoir	60
282	industrial	robot		60