Tailoring academic words to multidisciplinary EAP classes. Pojanapunya, P. (2019). TESOL Journal, 10(3), e00439.

The definitive version of this article was published as Pojanapunya, P. (2019). Tailoring academic words to multidisciplinary EAP classes. TESOL Journal, 10(3), e00439. It is available at https://onlinelibrary.wiley.com/doi/abs/10.1002/tesj.439

Tailoring academic words to multidisciplinary EAP classes

Abstract

Prioritizing vocabulary for instruction is central to English for Academic Purposes (EAP) research. Several word lists have been created to serve different needs, with many of them being restricted to a basic classification as general, academic, and technical words. While discipline-specific vocabulary lists are useful for particular disciplines at one end of a continuum, and more general, common core vocabulary lists are useful for a range of disciplines at the other end, this study identified multidisciplinary words that lie in the middle and could be used by disciplines at either end of the continuum. The study identified important words in corpora of research articles collected from six disciplines with regard to common disciplinary domains, in the hard and soft sciences. The target words of six disciplines were compared and examined with regards to whether they occur in only one, or many, disciplines. With consideration for variability of word use across disciplines within the same disciplinary domain, words in hard sciences showed higher variability than others. The findings are discussed in terms of academic word identification, selection and use in EAP classes.

Keywords: multidisciplinary EAP class, academic words, word list, keyword analysis, research article corpus

Introduction

Vocabulary knowledge has been recognized as crucial to academic success (Gardner & Davies, 2014; Masrai & Milton, 2018; Morris & Cobb, 2004). With the immense size of vocabulary in natural language, one of the main interests of research in English for Academic Purposes (EAP) is therefore to prioritize and select a target vocabulary to be taught in EAP classrooms (Coxhead, 2000; Durrant, 2014). Several word lists of limited vocabulary size have been created to correspond to vocabulary used in academic English.

Words can be classified into core words, core academic words, and technical words (Nation, 2001). First, core words are sets of high-frequency words that are assumed to be encountered by all students while engaging in common academic activities in a university. In addition to core words, core academic words often refer to words that are not core but are relatively frequent and appear in academic texts across multiple disciplines. Third, technical

words include highly discipline-specific words that require scientific knowledge to understand their meanings (Hyland & Tse, 2007; Paquot, 2010).

Word lists that are useful to all students regardless of their disciplines have been created over the years, e.g., Ghadessy (1979), Xue and Nation (1984), and Gardner and Davies (2014). The most well-known lists are West's (1953) General Service List (GSL) of 2,000 word families and Coxhead (2000)'s Academic Word List (AWL) of 570 academic word families. Moving towards more disciplinary specialization, several lists of specialized vocabulary targeted to specific disciplines have also been compiled, e.g., the nursing academic word list (Yang, 2015), the environmental academic word list (Liu & Han, 2015), the business word list (Hsu, 2011), and the engineering word list (Mudraya, 2006).

The creation and the usefulness of these word lists are likely to create uncertainty among teachers regarding whether EAP teaching should focus on discipline-specific academic words or common core words. Lists of specialized words would be useful for a particular discipline, and common core words would be useful to all students regardless of their discipline. This study argues that common core words needed for general EAP classes should be identified and presented as different sets of shared words between multiple disciplines that lie between the two ends of the continuum of word specificity, from specialized to common core words.

Prioritizing words to teach

A large number of word lists that are available for EAP classes have been used in two main ways. First, teachers have simply adopted the existing word lists in their class activities without any changes (see Paribakht & Webb, 2016 for applications of word lists). Second, teachers have adapted existing word lists, for example, by shortening an existing list through focusing only on the words appearing in target texts for class lessons. With word profiling tools, such as Range (Nation & Heatley, 2002) and AntWordProfiler (Anthony, 2014), teachers can do word profile analysis to compare a target text against any word list that can be imported into the program and used as a reference list to identify words in the text that are and are not covered in the reference list. The most frequently used reference lists are GSL (West, 1953) and AWL (Coxhead, 2000). Words in the target text that also appear in GSL are generally perceived as core words, whereas words found in AWL are interpreted as common academic words. Any words in the target text that do not appear in either GSL or AWL are assumed to be either technical or low-frequency words. Therefore, instead of adopting the existing word lists to teach, e.g., GSL or AWL, modified word lists are more selective and more specific to students' needs.

Existing word lists are useful resources for EAP teaching. However, one criticism of using a single general word list to teach students from a range of disciplines is that words in the general list are not really general, as some of them are used more frequently in only certain disciplines. For example, although words in the AWL are from a wide range of disciplines, some research suggests that the list could be more beneficial for some disciplinary domains compared to others, (e.g., Durrant, 2014; Hyland & Tse, 2007). More recently, the New Academic Vocabulary List (AVL) was created by Gardner and Davies (2014). The AVL is generally considered the better word list in terms of size and representativeness of the corpus resource as well as criteria for inclusion of words in the list. However, this list has also been criticized on account of the words in the list not being equally

used among different disciplines (Durrant, 2016). This imbalanced coverage of words across different disciplines is partly due to the fact that each individual discipline has its own conventions of language use. Even though the same words have been used across many disciplines, occurrences, collocational behaviour, and meanings of these words reveal differences among disciplines (Hyland, 2002; Hyland & Tse, 2007).

Furthermore, some of these existing general word lists may have been created based on false distinctions among general words, academic words, and technical words. Descriptions of these groups of words might lead to the conception that they are clearly separable. In fact, the boundaries between these groups of words are fuzzy (Mudraya, 2006; Paquot, 2010). For example, some words that are commonly used in general English could also be considered as specialized terms in a particular discipline. When common words appear with discipline-specific connotations, they often convey meaning beyond their general definitions, such as design, parameters, and model, all of which have specialized meanings in Statistics (Anderson-Cook, 2010). Therefore, the distinction among general, core academic, and specialized words based on the coverage of words in the AWL and GSL could be problematic, especially if teachers excessively focus on words from these lists. Words that are not covered in these lists and are often assumed to be technical words should not be neglected, because they might be technical words that are commonly used across multiple disciplines.

For these reasons, several scholars have suggested replacing separate lists of general service words, academic words, and technical terms with a single list that is either more specialized or has a larger common core vocabulary. For example, Ward (2009) built an engineering word list of 2,000 word families that contains both technical terms and general words. Gardner and Davies (2014) created the AVL as a common core vocabulary including many general high-frequency words that are notably frequent and often have specialized meanings in academic English. Likewise, the word lists created in this study are not based on the separation of general, academic, or technical words. Rather, each list includes words that are considered important academic words commonly used in multiple disciplines.

Creating word lists for multidisciplinary EAP classes

Most previous existing word lists have included either specialized or common core words. First, lists of specialized words are considered valuable for EAP classes for a specific discipline because the vocabulary is largely discipline-specific (Durrant, 2016). For example, a nursing academic word list was created specifically from research articles in nursing (Yang, 2015). Second, common core word lists have been created in an attempt to serve the needs of students from all disciplines. For example, the AVL was created based on a large corpus of academic English covering texts from a range of important academic disciplines to provide core academic words (Gardner & Davies, 2014).

Based on the idea of discipline specificity and the imbalanced coverage of common core words across different disciplines, word lists that lie between highly specialized and general words would be valuable as a supplementary resource.

This study aimed to identify words that are commonly used in two or more related or unrelated disciplines. Disciplines are regarded as related or unrelated based on a common classification of academic knowledge as "hard" disciplines, such as physical sciences, or "soft" disciplines, such as humanities (Becher & Trowler, 2001). This distinction classifies

microbiology and biotechnology as related disciplines since they are both in hard science. On the other hand, microbiology and sociology are considered as unrelated disciplines in this study since the former discipline is from hard science and the latter is from soft science. Several studies of EAP contexts have been conducted in the contexts of these boundaries (e.g., Gholami, Mosalli, & Nikou, 2012; Kashiha & Heng, 2014). In the context of course design, EAP courses can also be organized on the basis of these boundaries, i.e., courses for both hard and soft sciences that lie between the two ends of discipline-specific and general EAP courses.

This study is in relation to two main assertions. First, word lists should not be built on pre-existing lists or on the division of core words, core academic words, and specialized words. Rather, potentially important words for different disciplines should be extracted from texts that meet students' needs. Second, words to be extracted from the target texts should be potentially useful for (1) students from multiple disciplines under soft sciences or (2) students from multiple disciplines under hard sciences, as well as (3) students from both hard and soft science disciplines in multidisciplinary classrooms.

Consequently, words that are considered important across multiple disciplines within the same disciplinary domain are associated with a more specific EAP, whereas those that are considered important across multiple disciplines from different domains are associated with a more general class type.

Research purpose

At the intermediate position between discipline-specific and core vocabulary lists, this study identified words that are useful for multidisciplinary EAP classes from research articles in six different disciplines. Rather than using the prescribed norms for identifying specialized or common core words, this study employed keyword analysis (Scott, 1997; Scott & Tribble, 2006), a corpus-based method generally used to distinguish important words from one corpus to another due to disciplinary choices (Hyland, 2012), to identify important words of the six disciplines.

Although this study particularly focused on words for multiple disciplines (words found to be important in at least two disciplines), the number of specialized words found in a single discipline was also investigated because the number of these words could also roughly suggest the variability of words in hard and soft disciplinary domains.

The study

Corpora of research articles

Target corpora were research articles from refereed journals in six disciplines, i.e., biotechnology (BIO), mechanical engineering (ME), microbiology (MICRO), political sciences (POL), psychology (PSY), and sociology (SOCIO). These six disciplines were chosen because each of them is established as a distinct discipline, and two groups of the three disciplines were considered as the disciplines from the hard science and soft science domains. Each of the corpora comprised 80 research articles from international refereed journals in biotechnology (406,761 word tokens), mechanical engineering (481,312 word tokens), microbiology (394,439 word tokens) representing hard sciences, and political science (564,934 word tokens), psychology (380,719 word tokens), and sociology (703,798 word tokens) representing soft sciences.

Research articles were chosen as the data for this study for three main reasons. First, research articles are a key genre that is generally accepted as a crucial medium for disseminating disciplinary knowledge and as a concise version of academic texts highlighting special topics of the disciplines. Second, the reading and writing of scientific papers are fundamental concerns for many students who are required to conduct a special study (Liu & Han, 2015). Finally, using research articles as a source of words seems to correspond to the aim of EAP, "the teaching of English with the specific aim of helping learners to study, conduct research or teach in that language" (Flowerdew & Peacock, 2001, p. 8).

Keyword analysis method

In order to extract important words from texts from different disciplines, this study adopted a corpus-based methodology called 'keyword analysis' that identifies a word that occurs with a significantly high frequency in a target text or corpus when compared against a benchmark (Bondi & Scott, 2010). The keyword analysis method is known as a data-driven method identifying important information extracted from the given data itself (Paquot, 2010). Keywords can be identified using a log-likelihood (LL) statistic that is automatically computed by keyword generating programs, such as AntConc (Anthony, 2014a), and WordSmith Tools (Scott, 2015). The formula for computing LL can be found in Rayson (2008). Log-likelihood values report whether the words are identified as keywords by chance or because the words are frequently used in specific disciplines (Hyland, 2012; Rayson & Garside, 2000). Therefore, the method can highlight the vocabulary that is particularly salient to a given corpus.

To conduct keyword analysis, a benchmark is required to identify words that are important in a corpus of disciplinary texts. The British National Corpus (BNC), designed to represent general English, was used as a benchmark in this study. Keywords derived from a comparison between a corpus of each discipline (a target corpus) and the BNC are words that are significantly more frequent in a target corpus than in the BNC. In other words, they are statistically important words in a particular discipline when compared with general language.

Identifying and classifying important words for multidisciplinary practices

To identify and classify important words, or words that are commonly used in soft sciences, hard sciences, and both disciplinary domains, the three main stages employed in this study were: producing keyword lists (see Scott & Tribble, 2006), identifying keywords, and classifying the keywords into words commonly used in individual disciplines and disciplinary domains.

First, six keyword lists were generated by comparing each of the corpora against the BNC using KeyBNC (Graham, 2014), the open-access program for keyword analysis. Output from the program includes a list of keywords, frequencies of each keyword in the target corpus and in the BNC, and log-likelihood value for each keyword. Keywords with positive log-likelihood values are words that occur more frequently in a target corpus than in the BNC (Bondi & Scott, 2010).

Next, previous studies that used this method typically created a set of criteria that helped identify keywords that were likely to be meaningful for their research purposes.

Criteria used to identify keywords in this study helped identify words that occurred in many texts in the corpus and reduced the selection of rare or specialized words that were specific to fewer texts. The criteria include

- threshold (to have a manageable number of words for classroom teaching, potentially important words were the top 1,000 words of each of the six keyword lists);
- minimum frequency (to ignore rare words, keywords were those occurring at least five times in a corpus); and
- dispersion (to remove specialized words that appeared in only a few texts, keywords were those appearing in at least five percent of texts in a whole corpus).

The final process was to classify the candidate keywords into groups of words that were found in one keyword list or many keyword lists (both from multiple disciplines within the same disciplinary domain, and multiple disciplines from both domains). This study focused on keywords that were found in two or more keyword lists. The number of the keyword lists in which each keyword appeared was counted and presented as a ratio of the occurrences in the two major domains. For example, H3:S2 showed the occurrence of a keyword in five disciplines: three hard science (H) disciplines and two soft science (S) disciplines.

Table 1 shows how potential keywords were classified. In columns, keywords are specifically identified as keys in one discipline on one end and words that are shared across six disciplines on the other. In different rows, the shared words are also placed as words targeted to hard sciences, multiple domains, and soft sciences. For example, a word that was shared across three disciplines (3Ds) could be a word found in three disciplines under hard sciences (H3:S0), or two hard disciplines and one soft discipline (H2:S1). It would be placed among words targeted to hard disciplines. In contrast, a word found in three soft disciplines (H0:S3) would be placed among words targeted to soft disciplines.

Include Table 1 about here

Proportions of the keywords found as unique or shared words were calculated. The unique keywords were treated as domain-specific words, and shared keywords were treated as words for multiple academic domains. Coverage of words in GSL and AWL was also examined by using AntWordProfiler (Anthony, 2014b) to support explanation of the results.

Results and discussion

Rather than having a word list for a specific discipline on one end and a core academic word list for all disciplines on the other, this study further presents the word types at the intermediate position that are shared among two or more disciplines. Table 2 shows the number of keywords found in one to six keyword lists (1-6Ds), with the most common core keywords being those found in all six disciplines (H3:S3). They are also classified as keywords found in hard sciences (H), soft sciences (S), and multiple domains with ratios of disciplines in which keywords appeared. For example, from 566 keywords found in two disciplines (2Ds column), there are 252 keywords found in hard disciplines only (H2:S0 = 252), 244 keywords in soft disciplines only (H0:S2 = 244), and 70 keywords in both hard and soft disciplines (H1:S1 = 70). Keyword types are presented in Appendix A.

Include Table 2 about here

These groupings could provide more selective choices for teaching specific to more general EAP classes covering specific disciplines (each of the disciplines), multiple disciplines under the same domain, or multiple disciplines from both domains. For example, words in the H3:S3, H3:S2, H2:S3, H2:S2 groups could be useful in general EAP for multiple disciplines. Teachers might consider choosing words from H3:S0, H3:S1, or possibly H2:S1, H2:S0 for classes that are more specific to hard disciplines, and H1:S3, H0:S3, and possibly H1:S2, H0:S2 groups, for those more specific to soft disciplines.

Although the unique keywords (H1:S0 and H0:S1 groups) were not the focus of this study because there are a large number of existing specialized word lists available for specific disciplines, their high quantity shown in Table 2 (H1:S0 = 924, H0:S1 = 821) compared with other lists seems to support discipline-specific EAP courses, compared to broader ones.

Table 3 shows the percentage of keywords in each list that appear in only a single discipline, or two or more disciplines.

Include Table 3 about here

Approximately four to six percent of the keywords in each list are key in all six disciplines. This set of keywords could be useful for general EAP classes. Keywords found as shared keywords among the six disciplines are: (et) *al*, *analysis*, *analyzed*, *associated*, *based*, *compared*, *data*, *decrease*, *dependent*, *different*, *each*, *effect*, *effects*, *et* (al), *g*, *higher*, *low*, *measured*, *model*, *observed*, *related*, *relative*, *reported*, *results*, *sample*, *significant*, *significantly*, *similar*, *specific*, *study*, *using*, and *values*. These shared keywords are associated with

- research in general (e.g., *data*, *study*, *analysis*, *results*);
- data analysis (e.g., *measured*, *compared*, *analyzed*, *observed*); and
- reporting and comparing findings (*significant*, *significantly*, *reported*, *different*, *similar higher*, *low*, and *decrease*).

Teachers might use these vocabulary items in classroom activities, vocabulary materials, or assessment items for multidisciplinary classes (also see Donley & Reppen (2001); Hou (2014); Nushi & Jenabzadeh (2016) for ways of implementing vocabulary lists in classroom contexts).

Some of these keywords (e.g., *study, results, measured, observed, reported, different, higher, low*, *compared, decrease*) are also covered in the GSL, the most common 2,000 words in English that are typically assumed to have been mastered by undergraduate students before they take an EAP class. As some of these common words in GSL were identified as keywords in this study, the results suggest that these common words are important to certain disciplines. The findings suggest that teachers should not always rely on definite boundaries between common words and academic words with the assumption that common words in the GSL are not academic words. In fact, teachers might consider choosing these GSL words to teach in an EAP class for both their general definitions and their specific definitions within their disciplinary contexts. In a core EAP class for any discipline, teaching research-related vocabulary for enhancing the ability to read and write research, such as words related to data analysis and methods, could be useful.

Words that are shared across three to five disciplines (3-5Ds in Table A1) could also be useful for a multidisciplinary class. First, words that appear in four to five disciplines with a

roughly equal number of hard and soft sciences (H3:S2, H2:S3, H2:S2) can be grouped into two main categories that are

- words related to research (e.g., *conditions, experiment, experimental, samples, analyses, control, measurements, treatment,* and *error)*; and
- words related to describing, comparing and discussing findings (e.g., *increased, negative, positive, statistical, statistically, high, comparison, difference, lower, similarity, indicate, indicating, identified,* and *suggesting*).

Since these words are found in both hard and soft science disciplines, teaching these words using the contexts of the two disciplinary domains also gives students the opportunity to encounter variations of word use in different disciplines.

The following are keywords that appear in three to four keyword lists that are found in one disciplinary domain more often than the other (i.e., H3:S1, H3:S0, H2:S1 indicating greater occurrence of words in the hard science domain, and H1:S3, H0:S3, H1:S2 indicating greater occurrence in soft science). Some examples of hard science keywords are *calculated, concentration, induced, normalized, signal, solution, volume, column, culture, demonstrated, determine, morphology, properties, stress, characterized, formation, performed,* and *presence*. Some soft science keywords are *affect, bias, categories, centered, coded, conflict, dimension, disadvantaged, diverse, indicators, outcome, overall, composition, construct, diversity, dynamics, interactions, modeling, models,* and *multiple*.

Although these groups of words are not evenly distributed between hard and soft sciences, teachers might teach these items to expand students' vocabulary knowledge, e.g., on similarity and difference between the two domains. Some words, such as, *concentration*, *volume*, *column*, *culture*, *morphology*, *stress*, *coded*, and *composition*, have completely different senses in the hard and soft science disciplines.

By counting the number of sources of keywords, the results show that most keywords are found in one or two disciplines (1-2Ds), especially the keywords of biotechnology, mechanical engineering, and microbiology that are typically technical, e.g., *accumulation*, *antibodies*, *biomass*, *chromatography*, *conjugated*, *electrophoresis*, *hybridization*, *oxidative*, and *phenotype*.

Teachers might consider presenting these discipline-specific words to a class because the students might also encounter these words in their major subjects with different use and meaning. This input also provides a broad awareness of disciplinary variation. Some of the 2Ds vocabulary items could be useful for multidisciplinary classes. As can be seen in Table A1, 2Ds words include many pairs or sets of words with their derivatives and related words. For example, keywords appearing in two hard science disciplines are [*cloned, clones, cloning*], [*diluted, dilution, dilutions*], [*expressed, expressing, expression*], [*extracted, extraction, extracts*], [*induce, induction*], [*optimal, optimization*], [*produced, production, products*], [*quantification, quantified, quantitative*], and [*sequence, sequences, sequencing*]. Some keywords appearing in two soft science disciplines are [*contribute, contributions, contributors*], [*educated, education*], [*expect, expectations, expected*], [*explain, explanation, explanations, explanatory*], [*members, membership*], [*neighborhood, neighborhoods, neighbors*], and [*race, racial, racially*].

Teaching these words should help raise students' awareness of morphological variants and the use of these words in varied contexts.

These 1-2Ds words account for more than two-thirds of the words in the lists, especially in mechanical engineering (ME), which shows the highest proportion of unique keywords (66.32%) of any discipline. Overall, many keywords are shared among two or three disciplines, but a few of them are shared among three to six disciplines. The results support some previous research that argued that there is a need for a greater focus on disciplinary specificity in academic English courses and actually challenged the usefulness of general EAP courses because each individual discipline has its own way of using words (Hyland, 2002; Hyland & Tse, 2007). The findings seem to suggest that a discipline-specific course could be preferable to a course for multiple disciplines.

The previous section showed the number of words in specific disciplines as well as words that are shared among various disciplines covering hard and soft sciences. The following section presents the results when the six disciplines are considered as two major disciplinary domains. Proportions of the unique and shared keywords were calculated on the basis of their occurrence among the three disciplines under the same disciplinary domain. The proportions of the keywords used in hard and soft sciences are shown in Table 4.

Include Table 4 about here

Between the two disciplinary domains, there are more shared words among soft science disciplines than hard science disciplines. In each domain, the soft sciences have more words that are shared among political science (POL), psychology (PSY), and sociology (SOCIO) (28.32%, 34.60%, and 29.92%), while the hard sciences contain fewer words shared among biotechnology (BIO), mechanical engineering (ME), and microbiology (MICRO) (16.49%, 12.45%, and 14.57%). In other words, the results show more variability of words in hard than soft sciences. This suggests that multidisciplinary classes could be more applicable to soft sciences than hard sciences.

In contrast, discipline-specific classes are more suitable for hard sciences due to the specificity of words used in this disciplinary domain. However, there are still some words shared among three hard science disciplines, for example, *phase*, *decreased*, *calculated*, *concentration*, *induced*, *determined*, *normalized*, *resistance*, *substrate*, *fraction*, and *density*. For universities where an EAP class for a specific discipline is not practicable, these shared words, therefore, suggest that teaching multidisciplinary classes for hard sciences is still possible.

Because the results show a large number of unique keywords in each discipline, especially mechanical engineering (80.21%), the results support the idea that each academic discipline has its own conventions of language use (Hyland, 2002; Hyland & Tse, 2007). Collaboration between EAP teachers and subject teachers in teaching or raising students' awareness of these discipline-specific words might be particularly useful. EAP teachers are not required to teach specialised or discipline-unique words; however, sharing the list with subject teachers or teachers who teach discipline-specific content could be beneficial. On the one hand, EAP teachers could encourage students to be aware of discipline-unique words. On the other hand, the discipline-unique word list that is directly related to a certain subject could be of great use for classes instructed by subject teachers.

Conclusion

This final section provides a summary of major findings, brief comment on the pedagogical implications of the results, and some ways forward for further research. Overall, the current study proposes more options for choosing vocabulary in EAP classes, including multiple disciplines from the same and different disciplinary domains. The existing word lists can be viewed as two ends, one of which advocates the more restricted vocabulary teaching, i.e., discipline-specific, and the second that recognizes the existence of a common core academic vocabulary that can be taught to students in many disciplines.

Between hard and soft disciplines, the disciplines under the soft domain are more closely related in terms of shared vocabulary than the hard disciplines, each of which has more cases of unique vocabulary. This implies that a general course for multiple soft disciplines is likely to serve students' needs better than any similar course for the hard disciplines. Overall, the results show that the number of discipline-specific keywords is higher than the number of shared keywords, suggesting that, where applicable, discipline-specific EAP classes should be prioritized over general EAP ones.

The results also show many words that appear among multiple disciplines (the useful words suggested for practical use in the previous section are presented together in Table 5). The results suggest that setting up a course for students from many disciplines featuring either hard or soft sciences is still feasible. Furthermore, the results suggest research-related words that could be part of a common core academic vocabulary course. These research-related words include ones used for describing research background and methodology, especially words related to reporting and comparing results. Previous studies have suggested that particular words are not used in the same way across a range of disciplines (e.g., Durrant, 2014; Hyland & Tse, 2007). The use of these shared research-related words in different disciplinary contexts might be a potential issue for future investigation.

Include Table 5 about here

For multiple disciplines under the same domain, there are more shared words among soft sciences than hard sciences. Teachers of EAP for soft science disciplines could consider using the keywords identified by this study for their classes. However, there are more discipline-specific keywords in hard sciences. In this case, academic word lists for specific disciplines, both existing lists (e.g., Liu & Han, 2015; Yang, 2015) and specifically created lists as presented in this current study, could be useful. Teachers might choose the source texts for creating word lists based on their students' major subjects as well as learning objectives. For example, research articles and theses could be sources of words useful for writing, and textbooks and popular science articles could serve as sources of words that are useful for reading. Beyond the small set of keywords suggested in this study, students should be encouraged to learn vocabulary from texts and research in their respective disciplines.

The Author
Punjaporn Pojanapunya is a researcher at School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand. Her research interests include corpus linguistics, keyword analysis and corpus-based analysis of academic language. Her recent publication (with Watson Todd, R.) is Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis.

References

Anderson-Cook, C. M. (2010). Hidden jargon: Everyday words with meanings specific to statistics. In *Data and context in statistical education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.*

Anthony, L. (2014a). *AntConc* (Version 3.4.3) [Computer Software]. Tokyo: Waseda University.

Anthony, L. (2014b). *AntWordProfiler* (Version 1.4.0.0) [Software]. Available from http://www.laurenceanthony.net/software/antwordprofiler/

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: Society for Research into Higher Education Open University Press.

Bondi, M., & Scott, M. (2010). *Keyness in texts*. Amsterdam: John Benjamins.

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

Donley, K. K., & Reppen, R. (2001). Using corpus tools to highlight academic vocabulary in SCLT. *TESOL Journal, 10*(2□3), 7–12. https://doi.org/10.1002/j.1949-3533.2001.tb00027.x

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics, 35*(3), 328–356.

Durrant, P. (2016). To what extent is the academic vocabulary list relevant to university student writing? *English for Specific Purposes, 43*, 49-61. http://dx.doi.org/10.1016/j.esp.2016.01.004

Flowerdew, J. A., & Peacock, M. (2001). *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.

Gardner, D., & Davies, M. (2014). A new vocabulary word list. *Applied Linguistics, 35*(3), 305–327.

Ghadessy, P. (1979). Frequency counts, word lists, material preparation: A new approach. *English Teaching Forum, 17*, 24–27.

Gholami, J., Mosalli, Z., & Nikou, S. B. (2012). Lexical complexity and discourse markers in soft and hard science articles. *World Applied Sciences Journal, 17*(3), 368–374.

Graham, D. (2014). *KeyBNC* [Software]. Available from http://crs2.kmutt.ac.th/Key-BNC/

Hou, H. I. (2014). Teaching specialized vocabulary by integrating a corpus-based approach: Implications for ESP course design at the university level. *English Language Teaching, 7*(5), 26–37.

Hsu, W. (2011). A business word list for prospective EFL business postgraduates. *Asian ESP Journal, 7*(4), 63–99.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes, 21*(4): 17–29. http://dx.doi.org/10.1016/S0889-4906(01)00028-X

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing.* Ann Arbor: University of Michigan Press.

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse.* Cambridge: Cambridge University Press.

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

Kashiha, H., & Heng, C. S. (2014). Structural analysis of lexical bundles in university lectures of politics and chemistry. *International Journal of Applied Linguistics and English Literature, 3*(1), 224–230. http://dx.doi.org/10.7575/aiac.ijalel.v.3n.1p.224

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes, 39*, 1–11. http://dx.doi.org/10.1016/j.esp.2015.03.001

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44–57.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, *32*(1), 75–87.

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

Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes, 39*, 26–44.

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

Nation, I. S. P., & Heatley, A. (2002). *Range* [computer software]. Wellington: Victoria University of Wellington.

Nushi, M., & Jenabzadeh, H. (2016). Teaching and learning academic vocabulary. *California Linguistic Notes*, *40*(2), 51–70.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London & New York: Continuum.

Paribakht, T. S., & Webb, S. (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *Journal of English for Academic Purposes, 21*, 121–132. http://dx.doi.org/10.1016/j.jeap.2015.05.009

Rayson, P. (2008). *Log-likelihood and effect size calculator*. http://ucrel.lancs.ac.uk/llwizard.html (accessed 28 August 2018).

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora* (1-6). Hong Kong: Association for Computational Linguistics.

Scott, M. (1997). PC analysis of key words – and key key words. *System, 25*(2), 233–245. http://dx.doi.org/10.1016/S0346-251X(97)00011-0

Scott, M. (2015). *WordSmith Tools* (Version 6.0) [Computer Software]. Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* Amsterdam: John Benjamins.

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

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

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215–229.

Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes, 37*, 27–38. http://dx.doi.org/10.1016/j.esp.2014.05.003

Appendix A

Table A1

Lists of academic words for multidisciplinary EAP classes

	Hard sciences	Multiple domains	Soft sciences
6Ds		**H3:S3 (32 items)** (et) *al, analysis, analyzed, associated, based, compared, data, decrease, dependent, different, each, effect, effects, et* (al), *g, higher, low, measured, model, observed, related, relative, reported, results, sample, significant, significantly, similar, specific, study, using, values*	
5Ds		**H3:S2 (8 items)** *comparison, conditions, experiment, experimental, furthermore, of, respectively, samples* **H2:S3 (20 items)** *addition, analyses, control, differences, domains, e, factor, increased, indicate, indicating, levels, lower, mechanisms, negative, positive, potential, statistical, statistically, studies, test*	
4Ds	**H3:S1 (14 items)** *activation, characterized, deviation, evaluated, formation, method, n, obtained, parameters, performed, presence, shown, tested, used*	**H2:S2 (15 items)** *color, correlation, error, high, identified, increase, measurement, media, mediated, number, similarity, standard, suggesting, treatment, with*	**H1:S3 (33 items)** *also, average, behavior, behaviors, between, both, center, characteristics, coefficient, coefficients, conducted, consistent, contrast, correlations, difference, however, indicates, influence, interaction, interactions, level, modeling, models, multiple, predict, predicted, response, responses, sectional, theoretical, thus, variable, versus*
3Ds	**H3:S0 (41 items)** *b, c, calculated, carbon, cell, cellular, characterization, concentration, corresponding, decreased, density, determined, experiments, fig, fraction, generated, growth, induced, liquid, m, methods, morphology, normalized, oxidation, phase, profiles, resistance, resulting, signal, solution, strain, strains, subjected, substrate, surface, temperature, total, transformation, tubes, vector, volume* **H2:S1 (36 items)** *activity, anova, anti, approximately, biological, column, culture, cultures, cycles, demonstrated, determine, diffusion, domain, during, genetic, indicated, intensity, investigated, non, panel, pathways, process, properties, rate, ratio, revealed, selected, showed, signaling, stability, stress, target, type, were, whereas*		**H0:S3 (143 items)** *across, additional, affect, affiliation, among, Asian, assess, attachment, attitudes, baseline, behavioral, beliefs, bias, categories, category, causal, centered, coded, cognitive, comparing, conflict, context, contexts, controlling, correlated, covariates, degree, demographic, differ, dimension, dimensions, disadvantaged, discussion, diverse, empirical, empirically, engage, ethnic, evaluations, evidence, examine, examined, examining, extent, factors, favor, favorable, finally, finding, findings, focus, focused, gender, greater, group, groups, hypotheses, hypothesis, hypothesized, identities, immigrant, immigrants, impact, important, included, independent, indicators, individual, individuals, inequality, inferences, influences, interpersonal, items, labor, less, likelihood, likely, literature, may, measure, measures, more, moreover, negatively, organizational, our, outcome, outcomes, overall, parental, participants, participation, patterns, perceived, perceptions, performance, positively, predicting, predictor, predictors, predicts, preferences, previous, prior, questions, randomly, regarding, regression, regressions, relationship, relationships, research, respondents, role, salience, salient,*

			satisfaction, score, scores, settings, social, socioeconomic, specifically, standardized, status, strategies, stronger, strongly, suggest, suggests, support, than, their, these, toward, types, typically, understanding, variables, variance, whether, women **H1:S2 (34 items)** *abstract, analyze, and, are, clusters, composition, condition, construct, distributions, diversity, dynamics, function, identification, in, inclusion, increases, initial, larger, latent, longitudinal, magnitude, orientations, predictions, presented, processes, size, spatial, tests, therefore, this, traits, transition, variation, vs*
2Ds	**H2:S0 (252 items)** *absorbance, accumulation, acetate, acid, acids, actin, activated, active, agar, agarose, Aldrich, amino, amplification, amplified, antibodies, antibody, applied, assay, assayed, assays, bacteria, bacterial, beads, binding, bio, biomass, Biosystems, blot, broth, buffer, by, ca, cdna, cells, centrifugation, centrifuged, cfu, chromatography, cloned, clones, cloning, coated, collected, colonies, compounds, concentrations, confirmed, conjugated, conserved, containing, copper, cultured, curve, curves, deficient, degradation, described, detected, detection, digested, diluted, dilution, dilutions, dmem, dna, due, edta, efficiency, electrophoresis, eluted, emission, encoding, enzymatic, enzyme, enzymes, ethanol, exhibited, expressed, expressing, expression, extracted, extraction, extracts, fatty, flow, fluorescence, fluorescent, fold, forming, fractions, fragment, fungal, gel, gels, genbank, gene, genes, genome, genomic, gfp, glucose, glycerol, grown, h, harvested, homologous, host, hplc, hybridization, hydrolysis, ii, il, immune, inactivation, incubated, incubation, induce, induction, inhibition, inhibitor, inoculated, inoculation, intracellular, Invitrogen, isolated, kanamycin, kinetic, kit, l, lab, labeled, lipid, localization, magnesium, marker, markers, medium, melting, membrane, mg, mice, microbial, microorganisms, microscopy, mixture, ml, modified, mol, molecular, mouse, mrna, murine, mutant, mutants, nacl, nitrogen, nj, nm, nucleotide, nutrient, optimal, optimization, oxidative, pathogens, pathway, pbs, pcr, peptide, peptides, ph, phenotype, phosphate, phosphorylation, plant, plasma, plasmid, plasmids, plates, polymerase, previously, primer, primers, produced, production, products, promega, promoter, protein, proteins, protocol, purification, purified, qiagen, quantification, quantified, quantitative, reaction, reagent, recombinant, redox, reduction, region, residues, resin, resulted, resuspended, rna, rt, saline, sds, secreted, sequence, sequences, sequencing, serum, sigma, sodium, soluble, sp, stained, staining, sterile, sterilized, substrates, sucrose, supernatant, supplemented, surfaces, survival, suspension, t, temperatures, tissues, transcriptional, transfected, transformants, transient, treated, triplicate, usa, utilized, v, vectors, visualized, vitro, vivo, was, washed, wild, wt, yeast, yield*	**H1:S1 (70 items)** *ability, accuracy, amplitude, amplitudes, approach, authors, chain, cluster, clustered, component, components, conclusions, constructed, contact, content, criterion, cross, cycle, decreases, dispersion, disruption, distribution, dynamic, equal, equation, equilibrium, external, fatigue, fixation, frequency, increasing, internal, investigate, is, limitation, linear, linkage, linked, mechanism, modeled, neutral, note, observations, onset, orientation, oriented, parameter, points, pr, prediction, present, procedure, proportional, regime, sampling, sd, shows, sizes, structure, subgroup, threshold, tolerance, treatments, typical, value, varying, wave, weighted, yields, zero*	**H0:S2 (244 items)** *actors, additionally, adolescent, adolescents, advantaged, affects, age, American, americans, analyzing, argue, article, assigned, associations, attributes, attributions, benefits, biases, black, blacks, candidates, capita, capital, census, child, children, citizenship, civic, cohort, comparative, compliance, congruence, consequences, contextual, contribute, contributions, contributors, controls, counties, countries, covariate, cue, cues, cultural, dataset, depressive, descriptive, determinants, deviant, dichotomous, direct, discrimination, dummy, economic, educated, education, efficacy, efforts, elite, elites, emphasize, employee, employees, enforcement, engagement, errors, estimate, estimates, ethnicity, eu, evaluate, expect, expectations, expected, experience, experiences, explain, explanation, explanations, explanatory, exposure, favored, federal, feedback, female, focuses, for, frame, frames, framing, gdp, generally, geographic, global, health, heterogeneity, hierarchical, highly, Hispanic, Huber, hypothesize, identify, identity, implications, importance, include, income, index, indicator, indirect, individualistic, influenced, information, institutional, institutions, intention, intergroup, internet, interracial, interval, issues, job, judgments, knowledge, language, latinos, leadership, legislative, legislators, limitations, logistic, logit, macro, majority, maternal, mediating, members, membership, mental, microlevel, minorities, minority, mobilization, moderate, moderating, mood, mothers, motivated, multilevel, multinomial, multivariate, neighborhood, neighborhoods, neighbors, network, networks, nonsignificant, normative, norms, occupational, ols, online, opinion, organization, organizations, other, pattern, perceive, percent, percentage, perception, perspective, policies, policy, political, population, positions, practices, prejudice, primary, probabilities, probability, programs, propensity, proximity, public, race, racial, racially, random, ratings, reasoning, reflect, regardless, relatively, religious, researchers, resources, respondent, robust, roles, same, scale, scholars, self, ses, sex, significance, similarly, skills, socialization, societal, sources, spending, state, states, stereotype, stereotypes, strategy, students, substantively, supplemental, survey, surveys, table, tasks, tend, that, theories, theory, those, threat, ties, trait, turnout, u, vary, voter, voting, welfare, while, white, whites, workplace*

Table 1

Unique and shared keywords used in EAP classes for hard sciences, soft sciences, and both domains

Specificity	← Specific EAP			General EAP →		
#Disciplines	1D	2Ds	3Ds	4Ds	5Ds	6Ds
Hard sciences	H1:S0	H2:S0	H3:S0 H2:S1	H3:S1		
Multiple domains		*H1:S1*		*H2:S2*	*H3:S2* *H2:S3*	*H3:S3*
Soft sciences	H0:S1	H0:S2	H1:S2 H0:S3	H1:S3		

Table 2

Number of unique and shared keywords within and across disciplinary domains

Specificity	← Specific EAP			General EAP →		
#Disciplines	1D	2Ds	3Ds	4Ds	5Ds	6Ds
Hard sciences	H1:S0 (924)	H2:S0 (252)	H3:S0 (41) H2:S1 (36)	H3:S1 (14)		
Multiple domains		*H1:S1* *(70)*		*H2:S2* *(15)*	*H3:S2* *(8)* *H2:S3* *(20)*	*H3:S3* *(32)*
Soft sciences	H0:S1 (821)	H0:S2 (244)	H1:S2 (34) H0:S3 (143)	H1:S3 (33)		

Table 3
Proportions of keywords appearing in a single discipline or across disciplines

No	#Disciplines	1D (%)	2Ds (%)	3Ds (%)	4Ds (%)	5Ds (%)	6Ds (%)
1	Biotechnology	29.51	42.53	12.85	5.21	4.34	5.56
2	Mechanical engineering	66.32	10.62	10.75	6.29	1.83	4.19
3	Microbiology	38.19	38.04	10.89	4.14	3.83	4.91
4	Political sciences	36.89	28.45	21.86	5.59	3.23	3.98
5	Psychology	39.30	16.08	27.77	7.89	4.10	4.86
6	Sociology	34.78	29.27	22.70	6.04	3.02	4.20

Table 4
Proportions of keywords found in hard sciences and soft sciences that are unique words and shared words among two or three disciplines under the same disciplinary domain

Domains	Disciplines	Discipline-specific	Shared keywords		Total
		1D (%)	2Ds (%)	3Ds (%)	(%)
Hard sciences	Biotechnology	32.12	51.39	16.49	100
	Mechanical engineering	80.21	7.34	12.45	100
	Microbiology	40.64	44.79	14.57	100
Soft sciences	Political science	41.61	30.06	28.32	100
	Psychology	47.65	17.75	34.60	100
	Sociology	38.19	31.89	29.92	100

Table 5
Summary of the useful words for teaching multidisciplinary EAP classes

Domains	Shared keywords	
	3-4 disciplines (H3:S1, H3:S0, H2:S1)	2 disciplines (H2:S0)
Hard sciences	calculated, characterized, column, concentration, culture, demonstrated, determine, formation, induced, morphology, normalized, performed, presence, properties, signal, solution, stress, volume	[cloned, clones, cloning], [diluted, dilution, dilutions], [expressed, expressing, expression], [extracted, extraction, extracts], [induce, induction], [optimal, optimization], [produced, production, products], [quantification, quantified, quantitative], [sequence, sequences, sequencing]
Multiple disciplines	6 disciplines (H3:S3)	4-5 disciplines (H3:S2, H2:S3, H2:S2)
	(et) al, analysis, analyzed, associated, based, compared, data, decrease, dependent, different, each, effect, effects, et (al), g, higher, low, measured, model, observed, related, relative, reported, results, sample, significant, significantly, similar, specific, study, using, values	analyses, comparison, conditions, control, difference, error, experiment, experimental, high, identified, increased, indicate, indicating, lower, measurements, negative, positive, samples, similarity, statistical, statistically, suggesting, treatment
Soft sciences	3-4 disciplines (H1:S3, H0:S3, H1:S2)	2 disciplines (H0:S2)
	affect, bias, categories, centered, coded, composition, conflict, construct, dimension, disadvantaged, diverse, diversity, dynamics, indicators, interactions, modeling, models, multiple, outcome, overall	[contribute, contributions, contributors], [educated, education], [expect, expectations, expected], [explain, explanation, explanations, explanatory], [members, membership], [neighborhood, neighborhoods, neighbors], [race, racial, racially]