

## **A Corpus-driven Food Science and Technology Academic Word List**

**Rajab Esfandiari\***

*Assistant Professor of TEFL, Imam Khomeini International University,  
Iran*

**Ghodsieh Moein**

*M.A. student of TEFL, Islamic Azad University, Qazvin Branch*

**Received:** April 11, 2015; **Accepted:** December 23, 2015

### **Abstract**

The overarching goal of this study was to create a list of the most frequently occurring academic words in Food Science and Technology (FST). To this end, a 4,652,444-word corpus called Food Science and Technology Research Articles (FSTRA), which included 1,421 research articles (RAs) randomly selected from 38 journals across five sub-disciplines in FST, was developed. Frequency and range-based criteria were used to develop Food Science and Technology Academic Word list (FSTAWL). Word families had to occur in more than 19 journals, and they had to recur at least 134 times in the whole corpus. The computer programme RANGE was used to analyse the data. The results of frequency and range-based analysis showed that 1,090 academic words met the criteria of the study and constituted FSTAWL. The results also revealed that these words accounted for 13% of the coverage in the FSTRA. FSTAWL provides food science and technology non-native English learners who need to read a large number of RAs and to publish FST RAs in the English journals with a useful list of the most frequently used academic words, helping them to strengthen their academic reading and writing proficiency. The findings echo calls for creating more discipline-specific word lists to cater for the needs of specialized learner populations, providing implications for materials producers as well as explicit teaching of academic words.

**Keywords:** academic words, Food Science and Technology, Food Science and Technology Academic Word list, Range

---

**Authors' emails:** [esfandiari@hum.ikiu.ac.ir](mailto:esfandiari@hum.ikiu.ac.ir); [moein.q7@gmail.com](mailto:moein.q7@gmail.com)

## INTRODUCTION

Vocabulary learning is viewed as one of the most important components of learning another language due to its direct influence on language learners' reading and writing skills (Hirsh & Nation, 1992; Yang, 2015). The process of learning vocabulary has always been somewhat challenging in English for Academic Purposes (EAP) programs. The language learners' lexical knowledge—whether it is their general English knowledge or their specialized English knowledge—varies among the learners who are not usually proficient enough to deal with the requirements of their field (Hsu, 2013).

Vocabulary in EAP is particularly important for two main reasons. The first reason relates to the precious class time (Coxhead, 2013). EAP learners and EAP teachers should be cognizant of the very fact that what EAP learners in EAP classes do should be directly related to their lexical needs. The second reason has to do with group membership (Nation, 2011). “If learners are to become fully-fledged members of a particular community” (Coxhead, 2013, p. 116), they need to know the vocabulary the members of that community use.

Academic vocabulary causes a great deal of difficulty for learners since they are generally not as familiar with this group of words as they are with technical vocabulary in their own disciplines (Coxhead, 2000). Previous studies (cf., Young, 2015) have also indicated that such vocabulary knowledge is highly useful, and when it is combined with general vocabulary, it may cover more than 85% of academic texts. However, the most challenging questions regarding vocabulary learning in EAP settings is how many words, and which words, are to be taught in order to ease the comprehension of texts (Coxhead, 2000), considering the limited time and resources in EAP classes (Khani & Tazik 2013). In other words, not all words are equal in value for language learners, and some words need more attention (Wang, Liang, & Ge, 2008).

Researchers have developed several academic word lists (AWLs) to help language learners master academic words and set vocabulary goals for their independent study. However, to date, no study has focused on creating an academic word list (AWL) in Food Science and Technology (FST). The present study, therefore, was aimed at creating the first discipline-specific academic word list in Food Science and Technology

(FSTAWL). Furthermore, the study aimed to identify which academic words in FST would be found in Coxhead's (2000) AWL.

## LITERATURE REVIEW

In this section, four main types of vocabulary are explained to help readers better understand the nature of academic vocabulary. It is also attempted to summarize the findings of studies on academic vocabulary and AWLs, critically evaluate the merits and demerits of the findings, and finally discuss how they contribute to the present study.

### High-Frequency Vocabulary versus Low-Frequency Vocabulary

Nation (2001) identified four categories of vocabulary: (1) high-frequency vocabulary, (2) low-frequency vocabulary, (3) technical vocabulary, and (4) academic vocabulary. High-frequency words, as Coxhead (1998) put it, are those “essential to any learner of English” (p. 2). This category refers to words which occur frequently in any context and include words such as *it, can, also, from, these, they, some, over, had, and the*. Many of the content words such as *government, represent, adoption* are included in this category (Nation, 2001). One famous word list which is mainly based upon this category of words is West's (1953) General Service List (GSL). GSL is a long list of 2,000 word families, 80% of which are highly frequent.

Low-frequency words, on the other hand, as the name suggests, refer to those words with low dispersion and are rarely used (Nation, 2001). In Nation's (2011) words, the low-frequency words “consist of tens of thousands of words that occur very infrequently, are often restricted to certain subject areas, and thus do not deserve any substantial amount of classroom attention” (p. 531). Low-frequency words are “by far the biggest groups of words” (Nation, 2001, p. 12), including proper names, technical words of other disciplines, and words that are rarely used such as *pioneering, zoned, and aired* (Hsu, 2013; Yang, 2015). They cover approximately 5% of an academic text (Coxhead, 1998; Hsu, 2013; Nation, 2001; Yang, 2015).

## Technical Vocabulary versus Academic Vocabulary

Technical vocabulary refers to the words related to the topic and area of the text and differs from subject area to subject area (Nation, 2001). Some examples of technical vocabulary may include dyspnea, constipation, and tachycardia in Medicine; balance sheet, quantitative easing, stagflation in Economics; and refrigerant, atomize, blancher, toxicity, cryogenic in FST. Technical words provide a coverage of about 5% of the running words in a text (Nation, 2011). Hutchinson and Waters (1987) argued that technical vocabulary does not pose many problems for learners because they are “often internationally used or can be worked out from a knowledge of the subject matter and common word roots” (p. 166).

Also known as sub-technical vocabulary (Baker, 1988; Cowan, 1974; Flowerdew 1993), semi-technical vocabulary (Farrell, 1990), and academic words (Nation, 2001), academic vocabulary refers to “formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses” (Farrell, 1990, p. 11). Some examples of academic vocabulary include *use, sample, product, increase, value, low, measure, compare, follow, and perform*. These words are usually used to clarify academic notions (Liu & Han, 2015). These words belong to neither specific fields of study, nor to the general use (Coxhead, 1998; Hsu, 2013; Mudraya, 2006; Nation, 2001; Yang, 2015).

Providing a coverage of almost 9% of the running words in a text (Nation, 2001), language learners experience many difficulties when they use academic words since they are not so familiar with these items as they are with technical words (Yang, 2015). Much earlier, Anderson and Freebody (1981) reported similar problems, concluding that students most often identify academic words as unknown in an academic text. Also, these items do not occur in non-academic texts as frequently as they do in academic ones. Academic words almost always cause frustration among non-native learners since (1) the meaning of these words varies from one discipline to another, (2) it is common practice to use synonyms to refer to the same concept, and (3) these words have a Greco-Latin origin (Coxhead, 1998; Fahim, Fat’hi, & Nourzadeh, 2011). This has led many researchers, pioneered by West (1953), to make lists of the most frequently occurring words.

Researchers have grouped academic vocabulary into different categories. Baker (1988), for example, listed six categories of EAP vocabulary, including (1) lexical items expressing notions general to all specialized disciplines, (2) general lexical items bearing a specialized meaning in one or two disciplines, (3) specialized lexical items denoting different meanings in different disciplines, (4) general lexical items having limited meanings in different disciplines, (5) general lexical items used to describe, or comment on, technical processes, or functions, and (6) lexical items used to express the writer's intentions. One decade later, building on Baker's six categories, Dudley-Evans and St John (1998) offered two broad categories of academic vocabulary: "(a) general vocabulary that has a higher frequency in a specific field, and (b) general English words that have a specific meaning in certain disciplines" (p. 83).

Researchers use a wide variety of techniques to identify academic vocabulary. These methods have been documented in Chung and Nation (2003, 2004). Coxhead (2013) has neatly summarized these techniques which include "consultation with experts in a particular field, working with specialized dictionaries, developing rating scales, and using techniques from corpus linguistics" (p. 117). Corpus-based studies have proved promising in identifying academic vocabulary. As Coxhead rightly asserted, corpus studies "have been particularly useful for developing word lists for use in language classrooms and for independent study" (p. 118). Coxhead's assertion is confirmed by the increasing number of academic word lists being published in international journals and the empirical studies reported in the next section (cf., Lei & Liu, 2016).

The best known list of academic words developed through a corpus-based study is the AWL Coxhead (2000) created. The AWL includes 570 word families from a 3.5 words across four disciplines, including Humanities, Science, Commerce and Law. The word families in the list excluded West's (1953) 2,000 General words. Coxhead used frequency, range, and specialized occurrence to identify these words. Other researchers have followed her criteria and created more discipline-specific word lists since the publication of the AWL (see Coxhead, 2016 for an update).

In the following section, the findings from studies on academic word lists are summarized, the advantages, disadvantages, and criticisms

levelled at such word lists are examined, and the researchers who have compared their word lists with that of AWL are listed.

### **Previous Studies on Academic Word Lists**

In academic settings, through AWLs, the essential lexical items are provided to enhance proficiency (Yang, 2015). A number of AWLs have been developed by researchers in order to help students with the most useful, frequently occurring words across disciplines. Two of the earliest word lists were developed by Champion and Elley (1971) and Praninskas (1972), respectively, who developed their lists based on corpora of academic texts from a range of various university disciplines. However, one major shortcoming of these two lists was the corpora on which they were based were very small and represented only written language. Later, Lynn (1973) and Ghadessy (1979) adopted a different approach, basing their word lists on the annotations students made in their university textbooks.

Xue and Nation (1984) combined the previous four word lists (Champion & Elley, 1971; Lynn, 1973; Ghadessy, 1979; Praninskas, 1972) and put them into one word list called University Word List (UWL). UWL consists of about 836 words which were not in the West's 2,000 GSL, yet they were highly frequent in academic texts (Gardner & Davies, 2013; Khani & Tazik, 2013; Wang, Liang, & Ge, 2008). For a long time, UWL was assumed to be an ideal word list in developing the early versions of the two computer programs Range and Vocabulary Profile. However, this list, which was in fact an amalgam of the four compiled lists, lacked selection principles and shared some of the weaknesses of the above four lists. The word lists were based on small-sized corpora, and they did not represent a wide range of topics, genres, and text types. These lists were only based on written texts and did not represent spoken language (Coxhead, 1998, 2000; Gardner & Davies, 2013).

Although researchers developed some valuable AWLs prior to 2000, as the foregoing two paragraphs show, such lists did not necessarily meet lexical needs of language learners for some reasons. They represented written language, were based on very small-sized corpora, varied in data types, learner settings, selection criteria for identifying academic words, genre types, and corpus sizes, and used manual methods to identify the words due to the absence of computer

programs. These shortcomings prompted some researchers to develop more refined and more sophisticated methodologies (using larger corpora) to create more useful AWLs, as is shown in the following paragraphs.

The need for a list consistent with well-designed selection principles, based on a larger corpus of a wide range of topics of academic English, was strongly felt. Coxhead's (2000) AWL filled the gap. Using a corpus of 3.5 million running words, Coxhead applied frequency, range, and specialized occurrence to create the first AWL. Five hundred and seventy word families were identified. Out of her four sub-corpora, Commerce had the highest coverage (12.0%) while Science with 9.1% had the lowest coverage; Humanities and Law, the other two sub-corpora, fell in between. AWL covered 10% of its own corpus, and when combined with GSL, it accounted for approximately 90%, implying that out of 10 running words, language learners only find one word unknown (Coxhead & Nation, 2001).

Although Coxhead's AWL was a major breakthrough in developing the first comprehensive AWL which inspired many other studies, it was deeply flawed. One serious problem with her study was it did not include medical text types (Chen & Ge, 2007). The second major drawback was since the words in the 570-word list represented a wide range of disciplines, they did not necessarily convey the same meaning and occurred with different coverage rates across disciplines (Hyland & Tse, 2007). The other shortcomings included using written language to develop the corpus, employing a smaller number of shorter texts in Law compared to other sub-corpora, and incorporating an unequal number of disciplines in each corpus. These criticisms motivated other researchers to develop more discipline-specific word lists, as the following studies demonstrate.

Following Coxhead's AWL, a number of studies have been conducted, aiming to develop lists of the most frequent words for students of specific disciplines in order to meet their lexical needs. Mudraya (2006) developed a list of 1, 200 word families from a corpus of nearly two million running words for Engineering students. The list represented the most frequently occurring words Engineering students might need while reading or writing an academic text. Later, Wang, Liang, and Ge (2008) established a list of 623 non-GSL word families and called it Medical Academic List (MAWL). It provided a coverage of

12.24% in the 2-million-running-word corpus of the study and only 342, or 54% of the items, were found in AWL. These researchers primarily followed the Coxhead's three main selection criteria to develop their discipline-specific lists. However, Mudraya used only textbooks and Wang et al. only research articles to develop their relatively smaller corpora.

Hsu (2013), however, believed since more than half of the words in the MAWL were found in the AWL, this word list might be general in nature, not catering for lexical needs of medical ESP learners. This can lead to medical English learners' lack of exposure to enough discipline-specific words. Therefore, Hsu developed a far shorter and more manageable, yet more efficient AWL, for medical students. A list of 595 word families called Medical Word List (MWL) was established, of which only 76 words were present in the AWL. Vongpumivitch, Huang, and Chang (2009) created an Applied Linguistics Academic Word List, including 603 word families from a 1.5 million-word corpus called the Applied Linguistics Research Articles Corpus (ALC). The most recent work in this area belongs to that of Yang (2015) who developed a Nursing Academic Word List (NAWL). A Nursing Research Article Corpus (NRAC) of 1,006,934 running words was created through the collection of 252 nursing papers. NAWL consists of 676 word families and provides a coverage of 13.64% of the total NRAC. The interesting point is some academic words reported in the studies in the previous two paragraphs did not occur in Coxhead's AWL, confirming the fact that although academic words may occur across a wide variety of disciplines, some of them are definitely discipline-specific, unique to each discipline in terms of occurrence, use, and meaning.

Other than creating word lists in specific fields of study, some researchers have examined the coverage of their lists with those of Coxhead's AWL and West's GSL. Chen and Ge (2007) found that their Whole Paper Corpus (WPC) consisting of 50 English medical written research articles (RAs) with 190,425 running words covered 10.07% of AWL. Chen and Ge concluded that each section of a RA uses an appropriate number of academic words to achieve its purpose due to the fact that each section has its own focus. To find out the frequency, range, and the meaning of the AWL items across their corpus of 3.3 million running words, Hyland and Tse (2007) divided it into three sub-areas: Engineering, Sciences, and Social sciences. The findings showed that no lexical item in AWL occurred with the same frequency and



meaning across the disciplines. A few words of AWL occurred in all the disciplines, but in Biology Sub-corpus, AWL only accounted for 6.2% while for Computer Science Sub-corpus it was 16%. This, in fact, implies that the words were more useful for Computer Science students rather for Biology students. Therefore, Hyland and Tse concluded that since each discipline has its own unique way of expressing ideas and explaining matters, and its form of argumentation, word lists need to be more restricted and discipline-specific for academic students.

Valipouri and Nassaji (2013) conducted a study in an EFL setting to establish an AWL in Chemistry. A list of 1,400 word families was developed. The resultant list was called Chemistry Research Article Academic Word List (CRACL), and it was compared with the AWL and GSL. CRACL provided a coverage of 65.46% of GSL and 9.96% of AWL. These researchers did not use Coxhead's specialized occurrence because they claimed many general words "have sometimes different meanings, uses and collocations in specialized contexts" (p. 251). Valipouri and Nassaji acknowledged that their list consists of isolated items and cannot guarantee the knowledge of their use or meaning in Chemistry.

In some other studies, researchers have analyzed the meanings of academic words. Lam (2001) conducted an empirical study on Computer Sciences in order to detect the vocabulary problems that computer science students encounter while reading academic texts. She concluded that learners might be familiar with general words, yet they might not recognize the meaning of the very lexical item used in a technical context. Based on the fact that an academic word is semantically distinct from the same word in a general text, she suggested such terms be listed as glossary and available for those in the relevant area of the study. Martinez et al. (2009) developed an Agricultural Academic Word List containing 92 word families compiled from an agriculture corpus of 826, 416 running words, from research articles. They believed a list based solely on frequency would be of less use to learners than one based on pragmatic and semantic criteria, since the degree of topic relevance of the words relies on semantic association. This further confirms Hyland and Tse's (2007) claim that the same academic words may have different meanings in different disciplines.

As the above review shows, following the groundbreaking work of Coxhead, other researchers set out to develop more discipline-specific AWLs to help language learners in particular disciplines learn the most

frequently occurring academic words. More recent AWL studies, however, adopted a narrower approach to develop an AWL, using a single academic genre, a single discipline, a single mode of language to minimise the possible effects of genres, disciplines, and modes of language on resultant AWLs. Some studies also used general words to develop AWLs. Given the above review showing that general word lists may not help language learners achieve their goals in learning discipline-specific vocabulary, the need for developing FSTAWL is strongly felt to help ESP learners in FST learn academic words unique to this discipline. Building on only a single genre (RAs), a single discipline (FST), and written language, using frequency and ranged-based criteria, the present study filled this gap.

## **PURPOSE OF THE STUDY**

Providing students with a list of the most frequent academic words helps them to overcome the deficiency of the academic vocabulary competence. Although they might be proficient in general English, such word lists help them overcome the obstacles they face while reading a RA related to their field of study (Ward, 2009). FST students also need to read a number of books and RAs. In addition, in some cases, Master or Doctoral students are required to submit manuscripts in English. Discipline-specific word lists are very valuable for both teachers and students of the field, and an FST-based word list benefits learners in comprehending and writing an academic text. However, to date, no study has focused on FST to create an AWL. Therefore, this study seeks to establish the first Food Science and Technology Corpus and develop a Food Science and Technology AWL. The following research questions are addressed in this study:

1. What are the most frequent academic words in FST?
2. To what extent are the FSTAWL academic words used in the FSTRA?
3. How many academic words in the FSTAWL coincide with those in Coxhead's AWL and other word lists?

## **METHOD**

### **The Development of the Corpus**

In order to create the FSTRA, first, two content specialists were consulted to advise us on identifying the sub-disciplines of FST. They suggested choosing Food Chemistry, Food Engineering, Food Microbiology, and Food Technology. The list of these four subject areas was e-mailed to two other content specialists. They all agreed on the four sub-disciplines, but suggested Food Quality Control be regarded as another sub-discipline. These two specialists argued that evaluation is a core component of food processing and production, and it is common practice both in Iran and elsewhere in the world. They also explained that this sub-discipline has its own journals, research agenda, discourse communities, and conferences in the world. Next, with the help of the first two content specialists, journals related to each sub-discipline were identified. They suggested choosing journals with an Impact Factor (If) above 1.00, and those hosted by major international publishers, including Elsevier, Sage, Wiley, and Springer. In cases where IF was below 1.00, the content specialists recommended using them if they were hosted by international publishers and published RAs for at least 10 years.

Finally, a long list of 86 journals was established. The journals were categorized into the five sub-disciplines identified during the first stage. Eight journals for each sub-discipline except for Food Quality Control were randomly selected. For Food Quality Control, only six journals with an IF above 1.00 were identified, so all of them were included. The resultant list included 38 journals (Table 1) used to make the corpus in this study.

The RAs for the FSTRA were downloaded from the following two websites: <http://www.sciencedirect.com> and <http://www.freepaper.us/>. The RAs had to follow Introduction-Method-Results-Discussion (IMRD) format (Swales, 1990, 2004). Therefore, any RAs that did not follow IMRD format were eliminated. The selected RAs had to be published in the period spanning 2000 and 2014, and had to range in length between 1,800 and 7,000 words. Therefore, any RAs that did not follow the previous conditions were also eliminated. This left us with 1,421 RAs. Table 1 shows a breakdown of journals, RAs, and number of words in each sub-discipline.

**Table 1.** Journals, papers, and words in each sub-discipline

Sub-disciplines	No. of Journals	No. of Papers	No. of Words
Food Chemistry	8	294	979958
Food Microbiology	8	294	978444
Food Engineering	8	297	984604
Food Technology	8	284	975714
Food Quality Control	6	252	733724
Total	38	1,421	4,652,444

All the RAs were in PDF format which were first copied into Microsoft Word and later converted into text files so that they could be readable by RANGE computer program used to analyze the data. RANGE, downloadable at <http://www.victoria.ac.nz/lals/about/staff/paul-nation>, is a widely used computer program, developed by Heatley, Nation, and Coxhead (2002), which researchers use to create word lists. RANGE determines frequency and range of each word. Only Introduction, Materials and Method, Results, and Discussion sections were copied, but the Abstract, Conclusion, and Acknowledgments sections were left out. This was because RAs had to conform to the IMRD format.

### Word Selection Criteria

Word families were used as the unit of analysis, and Coxhead's (2000) criteria of range, frequency and specialized occurrence were followed to identify word families. The corpus from which AWL was extracted contained around 3.5 million words. For any word family to be included in AWL, it had to occur 100 times in the whole corpus, or 28.57 times per million words (pmw) and at least 10 times in each of the four sub-disciplines. The corpus of this study contains 4,652,444 words, so for each word family to be included in the list, the cut-off frequency should be at least 134 times in the whole corpus ( $4.7 \times 28.57 = 134.2 \sim 134$ ). Coxhead (2000) considered range as her primary and the most important criterion since any list created mainly based on the frequency can be biased by topic-related words and longer texts. Therefore, only word families that occurred in at least half of the 28 subject areas were included in the list. Following that, as for range, in this study a word family had to occur in at least half of the 38 journals—19 or more journals.

Coxhead used specialized occurrence to create AWL, as all the word families of the list had to be outside of West's (1953) General Service List (GSL), the first 2,000 most frequently occurring words of English. However, a major issue in developing field-specific word lists such as FSTAWL is whether to exclude GSL from the list. Some researchers have criticized the elimination of these words because "many general high-frequency words have a much higher frequency in academic English than in general English and often have special meanings in academic English" (Lei & Liu, 2016, p. 42). Billuroglu and Neufeld (2005), Ward (1999, 2009) and Valipouri and Nassaji (2013) ignored the distinction between general and academic words.

Therefore, current practice in creating discipline-specific word lists is to combine range, frequency, and general, high-frequency words. Following Coxhead (2000), Gardner and Davies (2013), and Lei and Liu (2016), Range and frequency were used, and specialized occurrence was excluded, to identify a more comprehensive list of academic words in FST. All the word families included in the final list met the following selection criteria:

1. Range: A word family had to occur in at least half of the journals—19 or more journals.
2. Frequency: A word family had to occur at least 134 times in the FSTRA.

## **Data Analysis**

As for data processing, standardization and normalization of the RAs were implemented. As for the standardization of the RAs, titles, figures, pictures, tables, charts, formulas, acknowledgments, reference lists, Bio data, appendices, authors' information, and some components in the text which the computer software would not be able to process were completely removed to eliminate any possible factors affecting the analysis of data and to ensure that the texts included in the corpus were readable by the computer program. In other words, what was copied was pure text.

Normalization of the words, on the other hand, was automatically done by RANGE (Heatley, Nation, & Coxhead 2002). RANGE reads all the derivations, or inflections, of a word as its basic form, or headword, and counts their range and frequency as one word family. For instance,

*accident, accidents, accidental and accidentally* are counted as one word by the computer program. Word family was defined by Bauer and Nation (1993) as the base word plus all its closely-related affixed forms. According to Coxhead (2000), “comprehending regularly inflected or derived members of a family does not require much more effort by learners if they know the base word and if they have control of basic word-building processes” (p. 218). This may clarify the reason for the adoption of word family in many word lists. In the end, all the RAs in each of the journals were copied and pasted into one text file named after its journal. This provided us with five sub-disciplines and eight text files containing all the RAs of the journals.

## RESULTS

### **The Most Frequent Academic Words in FSTRA**

To investigate the most frequent academic words in FSTRA, any word family had to occur in at least half of the 38 journals used in establishing the FSTRA and at least 134 times in the entire corpus. RANGE generated a voluminous output. Using this output, the words which did not meet the criteria for frequency and range as described in methodology section were eliminated. First, each word family repeated across at least 19 journals in FSTRA, and all word families occurred at least 134 times in FSTRA. Using these two criteria provided us with a preliminary list. Next, prepositions, pronouns, determiners, conjunctions, auxiliaries, particles, proper names, and acronyms were eliminated because they are not considered academic in the strict sense of the word. The remaining words constituted the word families of the list. This list included 1,090 academic words with a total frequency of 1,190,321.

*Use* with the frequency of 27,880 was the most frequently used word in the FSTAWL which occurred in all 38 journals while the least frequently used words were *silver, social, service, fan, and consent* which occurred 134 times in the entire FSTAWL in 31 journals. The words in the FSTAWL occurred in a wide range of journals in the FSTRA. Table 2 shows the 30 top most frequent academic words in the FSTAWL.

**Table 2.** The top 30 most frequent academic words in FSTAWL

Headwords	Range	Frequency
Use	38	27,880
Sample	38	19,064
High	38	16,008
Show	38	13,995
Difference	38	13,784
Product	38	12,862
Study	38	12,738
Increase	38	12,538
Temperature	38	11,935
Result	38	11,623
Analyze	38	11,005
Value	38	10,995
Effect	38	10,660
Water	38	10,077
Low	38	9,238
Active	38	8,869
Present	38	8,739
Significant	38	8,687
Time	38	8,626
Table	38	8,217
Concentrate	38	8,115
Add	38	8,040
Content	38	7,886
Treat	38	7,797
Process	38	7,606
Obtain	38	7,537
Extract	38	7,511
Determine	38	7,429
Food	38	7,227
Measure	38	7,227

As can be seen in Table 3, 616 (56.56%) out of 1,090 headwords occurred across 38 journals, and 474 (43.44%) occurred between 19 and 37 journals. For example, two words repeated at least in 19 journals, the minimum frequency cut-off point set in this paper.

**Table 3.** Journal coverage of academic words in the FSTAWL

No. of journals covered	No. of words	%
38	616	56.56
37	111	10.19
36	80	7.35
35	54	4.96
34	33	3.03
33	32	2.94
32	25	2.29
31	20	1.84
30	18	1.65
29	15	1.38
28	14	1.28
27	13	1.19
26	11	1.01
25	9	.83
24	9	.83
23	6	.55
22	7	.64
21	6	.55
20	8	.73
19	2	.18
Total	1,090	100

Table 4 lists the coverage of the words in the FSTRA. RANGE calculates tokens and types for a given corpus. A token refers to all occurrences of every word form in a corpus no matter how many times the same word form is repeated, but a type includes only different word forms in a corpus (see Schmitt, 2010). Percent figures in RANGE are based on tokens (also known as running words) than types. By implications, the number of tokens is always larger than the number of types. As can be seen, there are 4,652,444 tokens, 109,045 types, and 2,387 word families in the corpus. In total, 3,047,173 of the tokens were in the first and second GSL and made up 65.5% of the FSTRA.



Moreover, 409,499 of the tokens were in AWL that accounts for 8.80% of the total words in our corpus, and finally 1,195,772 of the words, accounting for 25.70% of the total FSTRA, were not included in any of the lists. Table 4 also shows that AWL and GSL cover 74.3% of the FSTRA, denoting a high coverage and suggesting that these two lists play an important role in FST.

**Table 4:** The coverage of the FSTRA in the base word lists implemented in RANGE

Word List	Tokens	%	Types	%	Word families	Coverage of FSTRA
1st GSL	2775565	59.66	3265	2.99	978	59.66%
2nd GSL	271608	5.84	2268	2.08	841	5.84%
AWL	409499	8.80	2428	2.23	568	8.80%
Not on the lists	1195772	25.70	101084	92.70	Not known*	25.70%
Total	4,652,444	100.00	109,045	100.00	2,387	100.00

\* The number was too high to be counted by the program.

### The FSTAWL Words Used in the FSTRA

Regarding the second research question, the FSTAWL was used as the base words in RANGE. As Table 5 shows, FSTAWL covered 13% of the FSTRA. In addition, the coverage of Coxhead's AWL and West's 1<sup>st</sup> and 2<sup>nd</sup> GSL in the FSTRA was determined as well, and the results are presented in Table 6. As Table 6 shows, the 2,000 most frequent word families of GSL accounted for 3,047,173 tokens, or 65.5% of the FSTRA.

**Table 5.** The coverage of FSTAWL in FSTRA

Word List	Tokens	%	Types	%
FSTAWL	604923	13.00	1081	0.99
Not in the list	4047521	87.00	107964*	99.01
Total	4,652,444	100.00	109,045	100.00

\* The number was too high to be counted by the program.

**Table 6:** The coverage of different base word lists over the FSTRA

Word lists	Tokens	coverage of tokens
1st GSL	2775565	59.66%
2nd GSL	271608	5.84%
AWL	409499	8.80%
Not in the list	1195772	25.70%
Total	4,652,444	100.00

Table 7 gives the coverage provided by GSL and AWL in the present research and in the studies conducted by Coxhead (2000), Martinez et al. (2009), Li and Qian (2010), Valipouri and Nassaji (2013), Khani and Tazik (2013), and Liu and Han (2015). AWL accounted for 8.80% of our FSTRA. This is lower than 9.06% coverage of AWL in Martinez et al.'s (2009) corpus of agricultural papers, or than 9.96% in Valipouri and Nassaji's (2013) chemistry corpus.

**Table 7:** Coverage of AWL and GSL in the present research and other studies

Studies	GSL (%)	AWL (%)	GSL + AWL (%)
Coxhead (2000)	76.1	10	86.1
Martinez et al., (2009)	67.53	9.06	76.59
Li and Qian (2010)	72.63	10.46	83.09
Khani and Tazik (2013)	76.4	11.96	88.00
Valipouri and Nassaji (2013)	65.46	9.96	75.42
Liu & Han (2015)	70.61	12.82	83.43
The present Study	65.5	8.80	74.3

The coverage of AWL in this study is also lower than other studies such as Li and Qian (2010) with 10.46%, Khani and Tazik (2013) with 11.96%, and Liu and Han (2015) with 12.82%. As for GSL, of the 2,000 words, only 740 occurred with a high frequency in the FSTRA. This suggests that almost a third of the GSL might be worth learning for FST students who need to read and, in some cases, write RAs. GSL accounted for 65.5 % of the corpus of the present study. This rate is lower than that of Martinez et al. (2009) with 67.53%, Li and Qian

(2010) with 72.63%, Khani and Tazik (2013) with 76.40%, and Liu and Han (2015) with 70.61%. The only exception is the study by Valipouri and Nassaji (2013) with 65.46% which is 0.04% lower than the GSL coverage in this study. This might be because of the size of the corpus as it is larger than those above or because of the genre chosen in this study. RAs were the focus of the present study while others worked on books, dissertations, newspapers or a combination of those. Word selection criteria could be another reason for low coverage rate in the present study. Most studies (e.g., Hsu, 2013; Khani & Tazik, 2013, Li & Qian, 2010) conducted on academic word lists followed Coxhead's (2000) three word selection criteria: range, frequency and specialized occurrence, while the present study excluded the third criterion. GSL and AWL provide a combined coverage of 74.3% in the FSTRA, which is 15.7% lower than the coverage reported by Coxhead and Nation (2001).

### **The number of words in the FSTAWL coinciding with those in AWL**

The procedure for answering the second research question was used to examine the third research question. The results showed that 350 academic words in FSTAWL were present in AWL. This accounted for 32.11% of AWL. Moreover, the coverage the FSTAWL and the AWL provided for each of the five FST sub-disciplines was examined, and the results were compared. Table 8 shows the results.

**Table 8.** Subject area coverage of 1,090 words in AWL and FSTAWL

Sub-discipline	AWL (%)	FSTAWL (%)
Food Chemistry	8.46	12.32
Food Engineering	9.07	15.25
Food Microbiology	8.54	10.79
Food Quality Control	9.63	13.67
Food Technology	8.51	13.13
FSTRA	8.80	13.00

## **DISCUSSION**

The present study investigated a 4,652,444-word corpus of 1,421 IMRD-format RAs in FST. The purpose was to identify the most frequently used academic words in FST RAs and to establish an AWL

for the students of the field. In addition, the coverage of the finalized list in the developed corpus and Coxhead's (2000) AWL was explored. Applying range and frequency criteria, 1,090 word families were identified. The words are ranked based on their frequency of occurrence across the whole corpus of the study. This illustrates the importance, priority, and the usefulness of the words compared to each other. A comparison of FSTAWL words with those of the AWL showed that many of the words in the AWL are not frequently used in FSTRA. Even those highly frequent words occurred with a different frequency of occurrence in these two lists, indicating that academic words are not used similarly across various subject areas. Moreover, the presence of some highly frequent non-AWL word families supports the need to establish field-specific word lists from texts and target genres that learners need to deal with in their academic disciplines (Hyland & Tse, 2007; Martinez et al., 2009; Valipouri & Nassaji, 2013; Wang et al., 2008).

The second finding of this study has to do with the coverage of the word list in our study and that of other studies. As Table 7 shows, the separate and combined coverage of these word lists in the present study is lower than the coverage reported by almost all the other studies. This low coverage may be due to the genre, discipline, corpus size, and selection criteria. Different researchers use different data types, corpus sizes, and different disciplines to create an AWL, all of which may affect the coverage rates. Most importantly, when corpus size is small, and frequency and range are set very low, the coverage rate is more likely to increase. For example, in Young's (2015) study, each word family had to occur only 33 times in the one-million nursing corpus and repeat in only 11 out of 21 subject areas. Applying these two criteria yielded in 677 word families. This large number of academic words certainly has a higher coverage in the corpus.

AWL coverage of FSTRA was lower than that of some other lists. The most likely reason is Coxhead used a variety of genres and a wide range of disciplines over a long time span. However, Martinez et al. and Valipouri and Nassaji used a single genre and a single discipline over a limited period of time. In the present study, a single genre and a single discipline were during a short time period. These three common features in these three studies may have accounted for a higher coverage rate of academic words in FSTRA. Therefore, it could be argued that the narrower the scope of the corpora, the higher the coverage of word lists.

The variety in the coverage rates shows that AWL and GSL differ in the role they play in different disciplines and exhibit different degrees of strength. In particular, the findings of the second research question and those of other researchers indicate that the words in the AWL are not equally useful for students of specific fields of study. In other words, since these two lists failed to account for 25.7% of the corpus of the present study, one-fourth of the words encountered by a student would be unknown to him/her. This stresses the need for learning the additional words provided by the FSTAWL.

The third finding of this study is concerned with the extent to which FSTAWL academic words are covered by AWL. Of the 1,090 words in the FSTAWL, 350 occurred in the AWL. This suggests the necessity for the list created in this study. When more general academic word lists such as AWL are used, not all words in such lists may serve the purposes students in fields such as FST have. Students can focus on those words in their own fields of study. The AWL contains 570 word families, 350 of which were found in the FSTAWL, meaning that 220 of the words on this list may not be needed by FST students and that learning them might be a waste of time and energy.

Furthermore, a comparison was made between the coverage given by the FSTAWL and the AWL in each of the five sub-disciplines. As presented in Table 8, in all the sub-disciplines, the lowest coverage AWL gave was in Food Chemistry (8.46%) while the lowest coverage of the FSTAWL was in Food Microbiology (10.79%). The highest coverage of AWL was in Food Quality Control (9.63%), whereas the FSTAWL provided the highest coverage in Food Engineering (15.25%). In total, the FSTAWL provided more coverage of the sub-disciplines than did the AWL. One possible explanation for such differences may be due to the general nature of AWL because AWL included four subject areas and many disciplines under each area. FSTAWL was, however, more specific, limited to a single parent discipline. The list created in this study provides a better coverage in the entire corpus and the sub-disciplines in this subject area. This means that by focusing on this list, students would learn more word families and would, therefore, be better equipped while writing or reading FST texts. The new academic words found in the FSTRA can direct learners with the most frequently occurring words in FST while reading or writing an academic text in the field.

## CONCLUSION AND IMPLICATIONS

This study showed the importance of creating a field-specific AWL. The results of corpus analyses showed that 1,090 word families were identified. The most and the least frequent academic words were also identified in the FSTAWL. The words in the FSTAWL were compared to those in AWL, but not all words in AWL were used in the FSTAWL. FSTAWL accounted for a lower coverage of GSL and AWL, compared to that of other studies. Out of the 1,090 word families of FSTAWL, only 350 occurred in the AWL.

The findings of the first research question also underscore the importance of more general, high-frequency words when academic word lists need to be developed. Excluding these general words may result in a lop-sided word list, distorting the nature of the academic words and offering an incomplete picture of the discipline, the most frequently occurring words of which are identified.

FSTAWL may have the following implications. FSTAWL is the first comprehensive AWL in FST and can be considered a pedagogically useful list which may help graduate students in FST to increase their knowledge of the most frequently used academic words. This list is an accessible and user-friendly list for researchers, teachers, and specifically students who need to equip themselves with the most frequently used academic words in FST papers. This way, they can be familiarized with these words, learn such academic words, and use them while reading or writing. However, learning the word lists alone will not suffice; they should be used in contexts as well. As Coxhead and Byrd (2007) believe, “academic success requires learning how to use academic vocabulary in writing as well as recognize it in reading” (p.143).

In addition, FSTAWL can be used for explicit teaching of vocabulary in EAP classes in which teachers know what needs to be taught and learners know what they need to learn (Hsu, 2013). The ability to consciously recognize the difference between the academic genres and their specific words from everyday conversation register can be achieved by attending these classes. These classes can help students be aware of the academic words and to identify them once they see them in an academic text such as RAs.

Yang (2015) stated that, “the concept of a word family is beneficial for learners because knowledge of a base word can facilitate the

understanding of its derived or inflected forms of words” (p. 36). This will expand the vocabulary knowledge of students with lower proficiency level.

Other than teachers and graduate students, material designers can benefit from this list incorporating the words into the academic reading and writing materials. Lists such as FSTAWL serve as a guide for ESP material designers in English for Food Science and Technology Purposes (EFSTP) curriculum preparation and English for Academic Purposes (EAP) textbook development. In fact, material designers can develop specifically designed academic English textbooks to teach FST academic vocabulary, and FST RA reading and writing, which in turn can effectively improve FST students’ proficiency in academic reading and writing.

This study has its own limitations. The first limitation was the nature of the corpus. Since spoken genre was not accessible, only the written genre was utilized. It is suggested that, in the future, researchers use written as well as spoken discourse to create word lists. The second limitation was that the functions of word families were not examined in FST. The third limitation relates to the relatively small size of the FSTRA. Although the corpus used in this study was comparatively larger than those in some previous studies, a larger corpus is needed to make the results more valid and reliable. Small discipline-specific corpora run the risk of distorting the number and frequency of academic words. In this study, four specialists were consulted to identify the main subject areas and journals of FST. Consulting a larger number of content specialists to seek their opinions on sub-disciplines and journals of a field of a study help researchers make more informed decisions about which sub-disciplines and journals to include in their corpora. The final limitation is concerned with the rationale for selecting RAs. In this study, abstracts and conclusions were not included in the corpus of the study. The absence of these two parts may have affected the number of academic words. In the future, researchers may consider using these two sections to create a word list.

### **Bio-data**

**Rajab Esfandiari** is an assistant professor at Imam Khomeini International University in Qazvin, Iran. His areas of interest include

teaching and assessing L2 writing, multifaceted Rasch measurement, L2 classroom assessment, and EAP and EAP testing.

**Ghodsieh Moein** is an M.A. in TEFL. She graduated from Azad Islamic University, Qazvin Branch. She is currently teaching conversation classes at language institutes in Qazvin, and her areas of interest include ESP and development of academic word lists.

## References

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4(2), 91-105.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Billuroglu, A., & Neufeld, S. (2005). The bare necessities in lexis: A new perspective on vocabulary profiling. Retrieved from [http://www.lextutor.ca/vp/tr/BNL\\_Rationale.doc](http://www.lextutor.ca/vp/tr/BNL_Rationale.doc).
- Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, 36(1), 1-23.
- Baker, M. (1998). Sub-technical vocabulary and the ESP teacher: An analysis of rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4(2), 91-105.
- Byrd, P., & Coxhead, A. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129-47.
- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5(5), 31-64. Retrieved from [http://faculty.edfac.usyd.edu.au/projects/usp\\_in\\_tesol/pastissues.htm](http://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pastissues.htm).
- Campion, M., & Elley, W. (1971). *An academic vocabulary list*. Wellington: New Zealand Council for Educational Research.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26(4), 502-514.
- Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15(2), 103-16.



- Chung, T., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251-63.
- Coxhead, A. (2016). Reflecting on Coxhead (2000). A new academic word list. *TESOL Quarterly*, 50(1), 181-185.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389-399.
- Coxhead, A. (1998). *An academic word list* (Unpublished Master's thesis). Victoria University of Wellington. Wellington, New Zealand.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge., & S. Starfield, (Eds.). *The handbook of English for specific purposes* (pp. 115-137). Chichester: Wiley-Blackwell.
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp.252-267). Cambridge: Cambridge University Press.
- Engels, L. K. (1968). The fallacy of word counts. *International Review of Applied Linguistics*, 6, 213-231.
- Fahim, M., Fat'hi, J., Nourzadeh, S. (2011). Wordlists in language teaching and learning research. *International Journal of Linguistics*, 3(1), 1-13.
- Farrell, P. (1990). A lexical analysis of the English of electronics and a study of semi-technical vocabulary (CLCS Occasional Paper No. 25). Dublin: Trinity College. (ERIC Document Reproduction Service No. ED332551). Retrieved from <http://www.files.eric.ed.gov/fulltext/ED332551.pdf>.
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231-244.
- Gardner, D. & Davies, M. (2013). A new academic vocabulary list. *Applied linguistics*, 35(4), 1-24.
- Heatley, A., Nation, P., & Coxhead, A. (2002). RANGE [computer software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, 17(4), 454-484.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge: Cambridge University Press.
- Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly*, 41(2), 235-253.

- Khani, R., & Tazik, K. (2013). Towards the Development of an Academic Word List for Applied Linguistics Research Articles. *RELC Journal*, 44(2), 209-232.
- Konstantakis, N. (2007). Creating a business word list for teaching business English. *ELIA*, 7, 79-102.
- Lam, J. (2001). *A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicography* (Research reports, Vol.3). Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22(1), 42-53.
- Li, Y., & Qian, D. D. (2010). Profiling the academic word list (AWL) in a financial corpus. *System*, 38(3), 402-411.
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39(3), 1-11.
- Lynn, R. W. (1973). Preparing word lists: A suggested method. *RELC Journal*, 4(1), 25-32.
- Martinez, I. A., Beck, S., & Panza, C. B. (2009). Academic vocabulary in agricultural research articles: a corpus-based study. *English for Specific Purposes*, 28(3), 183-198.
- Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purposes*, 25(2), 235-256.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, 44(4), 529-539.
- Nation, I. S. P., & Hwang K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- Nation, P. (2001). How good is your vocabulary program? *ESL Magazine* 4(3), 22-24.
- Praninskas, J. (1972). *American university word list*. London: Longman.
- Richards, J. (1974). Word lists: Problems and prospects. *RELC Journal*, 5(2), 69-84.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan: Chippenham and Eastbourne.
- Swales, J. M. (1990). *Genre analysis. English in academic and research setting*. Cambridge, England: Cambridge University Press.
- Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *English for specific purposes*, 12(4), 248-263.

- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33–41.
- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442–458.
- Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language*, 12(2), 309–323.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28(3), 170–182.
- West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.
- Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3(2) 93-242.
- Yang, M. (2015). A nursing academic word list. *English for specific purposes*, 37(1), 27-38.