Advances in Language and Literary Studies
Vol. 1 No. 2; July 2010

Prototype Software for Pedagogic Word List Harvest

Jayakaran Mukundan
Universiti Putra Malaysia

Ng Yu Jin
Universiti Tenaga Nasional, Malaysia

Hong Siaw Swin
Universiti Putra Malaysia

Vahid Nimehchisalem
Universiti Putra Malaysia

doi:10.7575/aiac.alls.v.1n.2p.256

Abstract

One of the important decisions to be made by English Language Teaching (ELT) material developers or educators in various disciplines concerns the selection of vocabulary items that a learner should learn intentionally or unintentionally. Learning new vocabulary from textbooks or academic texts should begin with the scrutiny of the frequent types of the vocabulary used in the texts and how they are used in terms of syntax or lexis. The development of various corpora and word lists for materials of an English for Academic Purposes (EAP) or English for Specific Purposes (ESP) programme ought to be regulated to ensure better learning of specialised vocabulary in a specific academic field. Vocabulary learning in EAP or ESP should always be well informed by the latest development of corpora in various fields which leads to the creation of useful word lists to socialise learners into their field-specific learning discourse communities. With the introduction of the word list creation prototype software, One-Stop Word List Creator 1.0 (OSWLC 1.0) will be the latest must-have asset for educators all around the globe. In this study, the flexibility and essence of the software will be discussed. OSWLC 1.0 is believed to facilitate better word list creation process for any given field in terms of user-friendliness and time efficiency.

Introduction

Problems with Textbooks and Vocabulary Learning in ELT

One of the important decisions to be made by English Language Teaching (ELT) material developers or educators concerns the selection of vocabulary items that a learner should learn intentionally or unintentionally. Research in the area of ELT material evaluation indicates that most materials are often developed in an ad hoc manner (Mukundan, 2009). Our previous research on the vocabulary load and distribution in Malaysian school language textbooks shows that they are commonly produced disregarding the fact that the choice of vocabulary should be made based on their frequency of use in word lists like West's (1953) General Word List (GSL) (Mukundan & Anealka, 2007, 2009). Research on the distribution patterns of grammatical items in Malaysian Secondary School English language textbooks also shows that these items have not been presented adequately in these textbooks (Mukundan & Roslim, 2009). Language textbooks should have been the ultimate tool used for language mastery, especially in learning of new vocabulary. Because of the shortcomings identified in the textbooks, the pedagogy of vocabulary is indeed affected in many ways.

Materials should be developed systematically based on needed essential words specifically to a field (see Menon & Mukundan, 2010). Research has shown that materials developed in the field of English for Sciences, are believed to

debilitate the pedagogic process of learning in terms of lexicon (Mukundan & Menon, 2007). One of the important questions is that what types of vocabulary educators should teach, when and how valuable it is to the classroom needs (Hirsh & Coxhead, 2009). Even early developments in English for Specific Purposes (ESP) sought to identify specific accessible essential vocabulary with segregation of various lexical features for ESP materials development (see Kennedy & Bolitho, 1984). Since most teachers are clueless when it comes teaching of the required types of vocabulary in a classroom, it is best if they are provided with a reliable word list to be incorporated into their pedagogical approaches. It is believed that most teachers and materials developers are not experts in word lists creation in specific fields when English language is concerned. Some teachers are also believed to be inexperienced in teaching vocabulary using word lists like the GSL and AWL.

According to Nation (1990; 2001), there are four types of vocabulary with majority of the academic words fall under the category of frequently used English words, followed by the academic, technical and low-frequency words. Frequently used English words are often linked to the General Service List (GSL) developed by West (1953) and it is one of the word lists which is still valid, despite being developed in the 1950s, till today (Coxhead, 2000; Nation, 2001). The GSL consists of the 2,000 most frequently used common English words in any academic texts. The next type of vocabulary would be the academic words. Before specific-field students master their specialised vocabulary in their respective fields, they should be exposed to academic vocabulary as it was developed over a wide range of numerous fields with significance coverage of many types of academic texts (Coxhead, 2000; Nation, 2001). One of the most prominent word lists for academic purposes is the Academic Word List (AWL) by Coxhead (2000). The list is made up of words that are supportive rather than dominating central meaning of a topic, covering 28 subject areas over the divisions of Commerce, Law, Science and Arts (Coxhead, 2000). A more detailed presentation of AWL will follow later since the framework of the present study follows as the processes used and criteria set in the development of the AWL.

The other type of essential specialised vocabulary in a subject area is the technical vocabulary. Technical vocabulary is a specialist domain of a system of subject knowledge which is closely related to a particular subject area (Chung & Nation, 2003; 2004). Technical vocabulary is essential specifically for learners with special purposes (Chung & Nation, 2003; 2004). Words from this category are "subject related, occur in a specialized domain, and part of a system of subject knowledge" (Chung & Nation, 2004, p.252). Vocabulary learning, particularly in the acquirement of technical terms of a specialised discipline, is important in students' comprehension of texts. Technical words should be made known to learners particularly for a specific field. However, there is little consensus on what technical vocabulary is. This is because there is no well established approach in determining which words are technical enough to be categorized under this grouping, and there are no studies to compare the effectiveness of approaches (Chung & Nation, 2003; 2004). Low-frequency words, on the other hand, are words which are low in coverage in academic texts, rarely used words or proper nouns.

According to Nation (2001), the level of technicality varies according to fields of study based on how limited the words are to a specific field. In this study, the definition of technicalness adopted is based on the definition provided by Menon and Mukundan (2010) as follows:

1. Highly technical words – these are words which appear rarely outside its particular field such as 'epithelial' and 'chromosome' in the science and medical fields.

2. Sub-technical words – these are 'context independent' words (Cowan, 1974, p. 391) which occur with high frequency across disciplines – academic vocabulary.

3. Semi-technical words – these are words which have one or more general English language meanings and which in technical contexts take on extended meanings.

4. Non-technical words – these are words which are common and have little specialization of meaning, for example 'hospital' and 'judge'.

Learning new vocabulary from textbooks or academic texts should begin with the scrutiny of the frequent types of the vocabulary used in the texts and how they are used in terms of syntax. Thus, a corpus of the target tests ought to be developed to understand the nature of the words better, especially the creation of a pedagogic corpus which is made up of academic texts designed for pedagogic purposes. The development of various corpora and word lists for materials of an EAP or ESP programme ought to be regulated to ensure better learning of specialised vocabulary in a specific academic field. Vocabulary learning in English for Academic Purposes (EAP) or English for Specific Purposes (ESP) should always be well informed by the latest development of corpora in various fields. The development of pedagogic corpus and word lists can contribute in ensuring more sophisticated pedagogy in

language in various contexts. In response to recognizing learners' needs in specific fields, extensive research has been conducted on identifying specialised vocabulary in academic texts which helps to socialize learners into their discourse communities (see Ward, 1999; 2009; Mudraya, 2006; Wang, Liang & Ge, 2008; Martínez, Beck & Panza, 2009).

A review of the available word lists in the literature

Corpus-based investigations are essential to elevate classroom instructions so that learners are exposed to authentic languages, where teachers' intuitions of essential words often mislead students (Reppen, 2001). The classic example of high frequency word list is the one introduced by Michael West in 1953 (Coxhead, 2000; Nation, 2001), which until now still provides significant coverage of running words in academic texts (Coxhead, 2000). High frequency words such as those included in West's (1953) General Service List (GSL) consisting of the most widely useful 2,000-word families, derived from almost 5 million running words in English, covering about almost 80% of most academic texts. This corpus was done manually over the period of years as computer sophistication was not available then but it serves as a comprehensive general word list for learners in English. Range, coverage and frequency have been the essence in developing word lists from corpora over the years (Nation & Kyongho, 1995).

Considerable effort then has been devoted to identify the vocabulary needed for academic purposes; for example the American University Word List (Praninskas, 1972); the University Word List (Xue & Nation, 1984); and the Academic Word List (Coxhead, 2000). However, the most influential word lists developed for use especially in tertiary institutions is the Academic Word List (AWL) developed by Coxhead (2000). This list has been widely cited and used for corpora analysis and comparison to determine essential academic words for specialised fields such as in the field of finance (Li & Qian, 2010), engineering (Ward, 2009), agriculture (Martínez et al., 2009), applied linguistics (Vongpumivitch, Huang & Chang, 2009), medical (Chen & Ge, 2007; Wang, Liang & Ge, 2008) and various disciplines (see Hyland & Tse, 2007). This word list was created from 414 academic texts incorporating 3,513,330 running words and 70,377 types of words in approximately 11,666 pages of text. The corpus was divided into four sub-corpora: arts, commerce, law, and science, each containing approximately 875,000 running words and each was further subdivided into seven subject areas.

The programme RANGE (Heatley & Nation, 1996), earlier version of RANGE (Heatley, Nation & Coxhead, 2002), was used for the analysis of the words in the corpus. The following criteria were considered in developing the word list:

> 1. Specialised occurrence: The word families included had to be outside the first 2,000 most frequently occurring words of English, as represented by West's GSL (1953).
>
> 2. Range: A member of a word family had to occur at least 10 times in each of the four main sections of the corpus and in 15 or more of the 28 subject areas.
>
> 3. Frequency: Members of a word family had to occur at least 100 times in the Academic Corpus.
>
> (Coxhead, 2000)

Coxhead and Hirsch (2007) would then create a pilot science-specific word list based on the specific research agenda of identifying a scientific word list outside the GSL and AWL as well as comparing the pilot list with the current AWL in terms of coverage in written scientific corpus. The basis of the pilot science word list development selection criteria was range, frequency and dispersion. They found out that there were still a lot of gaps to be filled in terms of the pilot written scientific word list as there exist reasonable frequency and range for words outside the GSL and AWL and if only larger scientific corpus were developed, the list could have been more significant (ibid).

Mudraya (2006) selected 13 recommended English-language textbooks in basic engineering disciplines which are compulsory for all engineering students regardless of their fields of specialization to form her corpus. After the texts were digitized, the materials formed a corpus of nearly 2 million running words and more than 18,000 types. After analysis using WordSmith Tools software and comparison with the British National Corpus (BNC) and COBUILD Bank of English Corpus, the 1260 most frequent word families of 8850 words were included in the Student Engineering Word List based on the frequency of the three corpora.

Also, Ward (2009) developed a word list, the Basic Engineering List (BEL), using foundation engineering textbooks on topics such as materials in engineering, statics, fluid dynamics, and basic thermodynamics from various

engineering sub-disciplines (chemical, civil, electrical, industrial and mechanical). He consulted lecturers in five engineering faculties for the recommendation of books based on their expertise. About 271,000 words were randomly selected from 25 textbooks (five from each sub-discipline) with approximately 10,000 from each book. The word count and discrimination were carried out using RANGE software (Heatley, Nation & Coxhead, 2002).

> Engineering Corpus (EC) is representative and at least reasonably comprehensive (representing a range of topics in a variety of major engineering fields), balanced (giving equal importance to each field), genre-specific (only textbooks are represented) and relevant to student needs (textbooks for later years of undergraduate study) (Ward, 2009)

Ward (2009) believes that the word list BEL is able to provide wider coverage of engineering sub-discipline words as Coxhead's academic corpus does not include words specific to engineering. Thus, a 299-word Basic Engineering List (BEL) (Ward, 2009) was developed to aid students with low proficiency in English but who still needed the language of engineering for smooth transition into the discourse community of engineers. The word list BEL supports the notion that word lists developed from specific-field corpora give a much better coverage than general academic word lists. This is also indicated in his previous research on an engineering word list (EngList) (Ward, 1999).

The Medical Academic Word List (MAWL) (Wang, Liang & Ge, 2008) was developed from a specialised corpus containing 1,093,011 running words from 288 written texts of research articles in English for Medical Purposes (EMP), downloaded from the database ScienceDirect Online (http://www.sciencedirect.com). A total of 32 subject areas texts were sampled with three journals from each of the 32 subject areas. Articles from 96 selected issues were sorted according to the Introduction-Methodology-Result-Discussion (IMRD) format. In addition, articles which were not written by native English speaking writers or shorter than 2000 running words or longer than 12000, running words were eliminated. Their criteria (adapted from Coxhead, 2000) of the selection of words to be included in the MAWL were as follows:

> 1. Specialised occurrence: The word families included had to be outside the first 2000 most frequently occurring words of English, as represented by West's GSL (1953).
>
> 2. Range: Members of a word family had to occur at least in 16 or more of the 32 subject areas.
>
> 3. Frequency: Members of a word family had to occur at least 30 times in the corpus of medical research articles. (Wang, Liang & Ge, 2008)

Wang, Liang and Ge (2008) consulted two experienced professors in the field to assist them to select the essential words to be included. The final list of the MAWL consists of 623 word families, which accounts for 12.24% coverage of the medical texts. It is then deemed important to have a field-specific or specialist word list which takes into consideration words that are found from the core textbooks in a programme and one that provides a better coverage than any other academic word list (Mudraya, 2006; Ward, 2009).

Word List Creation

The creation of word lists is indeed beneficial to both EAP and ESP to socialise learners into their field-specific learning discourse communities. However, the creation process may take a long time and it requires patience and expertise from both the researchers and the teachers who are in need of technical word lists pedagogical purposes. The digitization process can be tedious and the reliability of the analysis of the edited texts can sometimes be questionable as the laborious steps can sometimes be overlooked. Human factor can be one of the major doubtful reliability when creating a word list as different researchers will consider different parameters in their analysis. From all the procedures discussed above, it is apparent that the basis of creating a word list of academic or specific in nature can be divided into three major parts. The frequency of occurrence, the range of occurrence and the specialised characteristics of the target vocabulary or word list are the three essential criteria in word list development. Xue and Nation (1984) created their University Word List (UWL) based on similar grounds which lead to the development of the comprehensively designed criteria of the development of the AWL. Conceptually, the main framework towards the development of this software came from the detailed method proposed by Coxhead (2000) in developing the AWL, using RANGE (Heatley, Nation & Coxhead, 2002). A new hybrid of software will

emerge from this study to solve the discrepancies and reliability issues in the current available concordance software. Thus, in this research, the new and innovative software, incorporating the essential word list development criteria will be introduced to the world of EAP and ESP in the ESL and EFL context. It is called the One-Stop Word List Creator version 1.0 (OSWLC 1.0), which is believed to be user-friendly and efficient in harvesting customized or specialised word list for different needs. The development of specialised vocabulary list is the main goal in the present study by using and testing this newly developed software programme. The word list created by the software programme will then be compared with a semi-manually developed word list using WordSmith 5.0 (Scott, 2010). The software programme will be elaborated in the results and discussion part and compared with similar software in the field in terms of its user-friendliness and simulation speed.

Preview of One-Stop Word List Creator 1.0 (OSWLC 1.0)

One-Stop Word List Creator 1.0 (OSWLC) is created using Microsoft Visual C# general purpose programming language assisted with Microsoft .NET framework 2.0. Current Microsoft.NET framework is at version 4.0 while the used version was 2.0 is due to the dependency issues. Microsft.NET framework 2.0 is incorporated in most of the windows that after windows XP SP1 while Microsoft.NET framework 4.0 is not even provided in the latest Microsoft Operating System which is Windows 7. Microsoft .NET framework 4.0 requires manual download directly from the official Microsoft website. In other words, this program is only dependent on the operating system of Windows XP with Microsoft.NET framework of at least version 2.0. Microsoft Visual C# was chosen as the programming language due to its 'extreme features' on items identification, in this case target words identification based on threshold input or limits. Every item in this program such as a book, a folder (category), or a word list are all assumed to be a real object and created as a virtual object in terms of visualization as well as in codes (programming). The word analysis concept used in this program is divided into a few parts, splitting strings, sorting strings, and counting strings. All parts are performed using methods that directly access through assembly level which result in speedy processing. All images used in the program are converted into controls and all the 'drag and drop' are updated in real time matter.

The text system for this program is Unicode instead of ASCII, which means it can be easily converted into a multiple language support tool like AntWordProfile (Anthony, 2009) which can process text in Unicode. However, for now it only focuses on analyzing English lexis only and the Unicode system is prepared for future version with Multi-lingual support. Also, this software is more 'word sensitive' as compared to those of other similar software programmes. For example, a word like 'Dysfunction1' is not counted as a word or headword of "Dysfunction" in RANGE (Heatley, Nation & Coxhead, 2002) and WordSmith 5.0 (Scott, 2010). In fact, the exclusion of such a word can affect the frequency count of the particular word in a corpus.

The selected literature reviewed in this study suggests the need of a new model of word list creation software in ESP/EAP context so that the word list creation process can be better facilitated or effortlessly executed. Curriculum or syllabus developer and even teachers or researchers would need a software program like this to create word lists for effective pedagogic purposes. The objective of this study aims to introduce and highlight the major functions of the newly developed software which acts as a basic tool for word sifting, categorization and word list harvest. The next goal would be testing and validating the software by comparing the output with a rigorously 'semi-manual' developed word list in the field of nursing (sample corpus created for testing purposes).

The research questions are as follows:

1) What would be the main functions of the new model of the word list creation software?

2) To what extent do the words differ from a manually created nursing word list?

The development procedure of WLC

The major highlights of the functions of the software will be illustrated in the findings with selected figures. In order to evaluate the reliability of the software for the purpose of this study, the text files of major nursing textbooks from a pedagogic nursing corpus (sample corpus) was obtained for the pilot testing. For the purposes of adapting the range, frequency and representativeness criteria in word list development, the 'simulation' process of word list creation carried out using the created sample corpus. The texts were then divided into two categories according to the original textbooks; specialised technical nursing education textbooks which are very technical in nature and semi-technical nursing education textbooks to demonstrate the 'range' setting in this programme.

The nursing word list created with the One-Stop Word List Creator 1.0 and the'semi-manual' method using a 'peer software' of WordSmith 5.0 (Scott, 2010) will be discriminated in terms of its similarity and differences(if any)

adopting the same criteria discussed to assess the validity and reliability of the target software. The word list production process will also be compared in terms of time economy, whether which tools adopted will take less processing time to harvest the word list and its user-friendliness. Researchers and educators will definitely prefer software which requires minimal technical knowledge and swift in obtaining the output. In this study, the word list developed by all the tools will simply be named as customized Nursing Word List (NWL).

Lexical processing and word selection criteria

The word selection criteria are generally similar to those of the AWL (Coxhead, 2000) and the MAWL (Wang, Liang & Ge, 2008), but with some variations. The NWL features academically technical or semi-technical nursing education words which have fulfilled the range, specialised lexical items and frequency criteria which are similar in procedure to both the AWL and MAWL criteria for inclusion into those lists. Technical or semi-technical nursing education words are also difficult to guess from context as student nurses are yet to be expert practitioners while studying in universities or colleges. Without a good background in that technical area (learning technical words is closely related with learning the subject), these technical words are hard to comprehend or guess, especially the terminological words which have very narrow range (Nation, 2001). In summary, the criteria set for the inclusion of the specialised nursing education words are:

I. The words must be found in any three out of five of the technical nursing education textbooks or in both the non-technical nursing education textbooks with at least 10 appearances in each book. This criterion addresses the range of the academic nursing words.

II. It must be a specialised list exclusively for the nursing field. Thus, the specialised lexical items are words that are found outside the GSL and AWL, categorized as academic words (Nation, 2001).

III. The specialised words must occur at least 100 times in the texts to be included in the list. Academic words are significant in the field if the frequency of occurrence is evidently high (Ward, 1999; Mudraya, 2006) considering the corpus is as large as Coxhead's AWL.

Results and Discussion: The functions within One-Stop Word List Creator 1.0 (OWLC 1.0)

The software was designed to be as user-friendly and efficient as possible. The 'work flow' or main functions of this programme are shown in Figure 1 and its characteristics in Table 1.

Table 1: Characteristics of Functions in OSWLC 1.0

Item	Function
Make new book(s)	File loader in text file format (.txt)
Set general filter	Frequency and range filter/settings
Process book(s) input	Word Loading and Distribution Patterns(computerized concordance analyzer)
Generate wordlist	Word list generation with words discrimination functions from that of the GSL and/or AWL

Figure 1: Main Interface of One-Stop Word List Creator 1.0 (OSWLC 1.0)

The best part of this software is that it leads the users step by step in the word list creation process. Users must complete keying in the required data or entry before the next stage can be processed. The indicator in this software will be lighted up should the users are in a particular step. The following illustrations will demonstrate the steps or 'flow' stages involved before a word list can be harvested. Figure 3 shows the addition or loading of files into the programme.

Figure 2: The Loading of Text Files into OWLC 1.0

After the files are loaded, users should click on the 'next step' button. Figure 4 illustrates the following steps where the corpus or 'books' were formed

Figure 3: Pre-Filter Settings Process

Users can now set the required filter frequency of a particular word appearing in the overall corpus and in every individual book. Figure 4 shows one part of the criteria discussed above where a word should appear in any THREE of the technical nursing textbooks with at least 10 words occurrence. In addition, the particular word must appear at least 100 times in the corpus as well. Users can change the settings anytime based on their requirements and the size of the target corpus.

Figure 4: Filter Settings in OSWLC 1.0

Users at this stage are required to click on the 'next step' button again. The next step involves the analysis of the text files. The frequency and distributions of the words across the textbooks are made visible at this stage and it is illustrated in Figure 5.

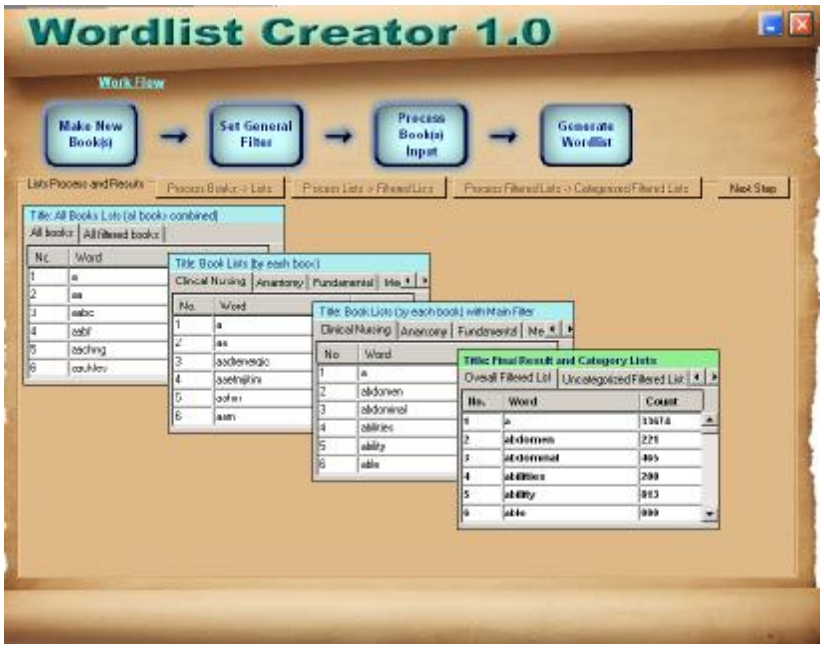


Figure 5: Words Loading and Distribution in OWLC 1.0

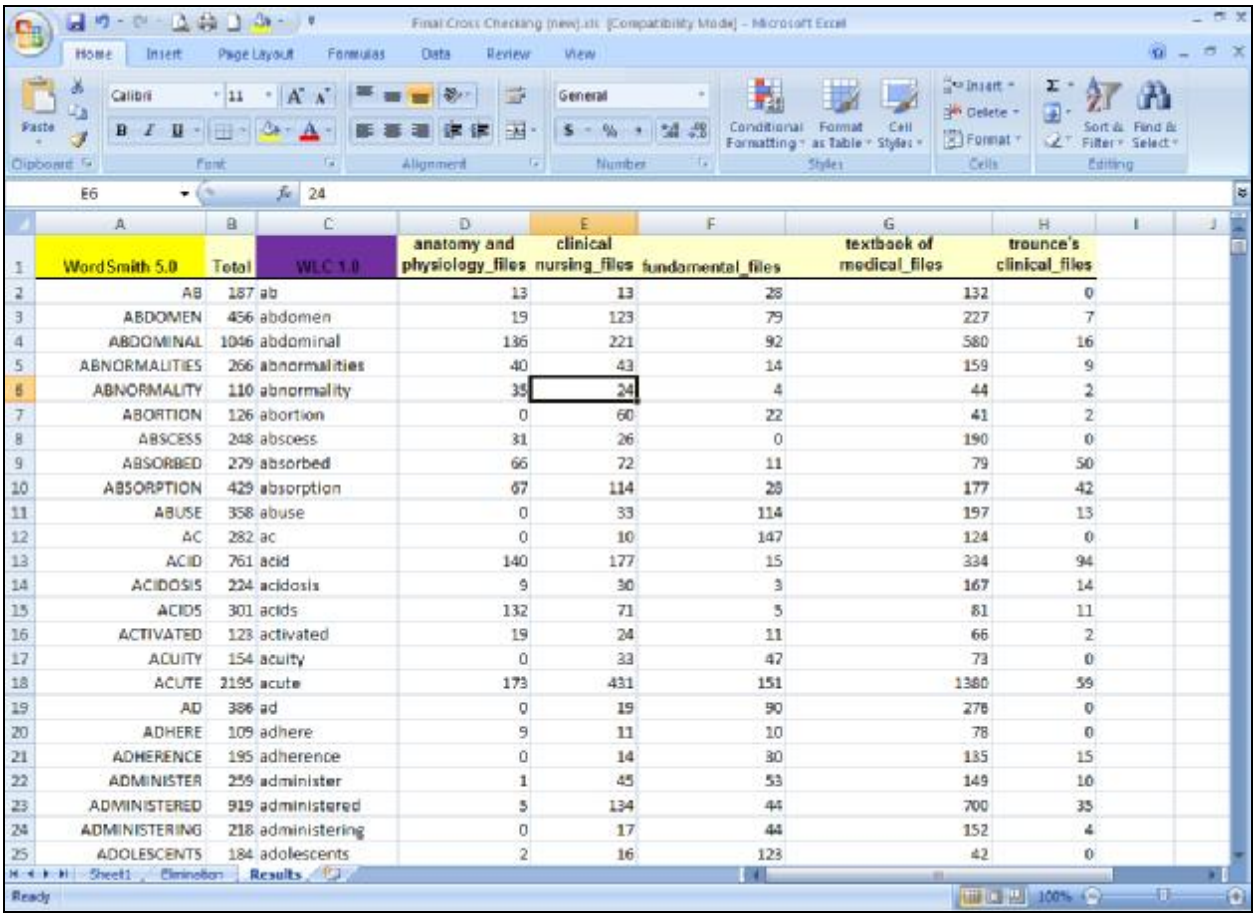
In the word list generation stage, users will be asked whether the similar words from the GSL and/or AWL of the target words be removed. If users want to activate this function, the 'add word list' function must be clicked. Figure 6 shows that the GSL and AWL filters were used and the word list size is limited to 3,000 words only. Users can change the settings of the size of words in the word list created freely.

Figure 6: Pre-Word List Generation Step in OWLC 1.0

The moment the button, 'generate complete wordlist' is clicked, the final word list, according to the desired criteria set, is developed. From the initial stage to the last stage, users must be aware of the criteria which are desired based on previous corpus linguistics recommendations or development. It is indeed an effortless process, following a string of user-friendly platform, to develop a word list for any fields of study. The word list formed can also be saved into text file format easily. Figure 7 illustrates a sample of the output from the word list generation from this programme. All the diagrams discussed are the elaborated answer to research question 1 (*What would be the main functions of the new model of the word list creation software?).*

Figure 7: The Word List Generation Process Completed

In order to answer research question 2 (*To what extent do the words differ from a manually created nursing word list?*), extended research was carried out to find out the efficiency in terms of time taken to run the software. Also, the study is aimed to determine the extent of discrepancy (if any) of the word lists created with the prototype and semi-manual approach. A direct approach of one to one comparison was adopted; a fully 'automated' approach using WLC 1.0 versus a 'semi-manually' approach of using WordSmith 5.0 to generate the word list based on the criterion discussed above. The time taken to generate the word list was recorded as well. The results showed a total of 913 words generated from both WLC 1.0 and WordSmith 5.0 based upon the criteria set of frequency, range and specialised. The time taken for WLC 1.0 to perform the entire analysis is approximately 15 minutes. However, for the 'semi-manual' approach, it took a total of approximately 60 minutes to filter all the words using WordSmith 5.0 and Microsoft Excel as platform, provided that human error is negligible. Figure 9 shows the excerpt of the results obtained.



	WordSmith 5.0	Total	WLC 1.0	anatomy and physiology_files	clinical nursing_files	fundamental_files	textbook of medical_files	trounce's clinical_files
2	AB	187	ab	13	13	28	132	0
3	ABDOMEN	456	abdomen	19	123	79	227	7
4	ABDOMINAL	1046	abdominal	136	221	92	580	16
5	ABNORMALITIES	266	abnormalities	40	43	14	159	9
6	ABNORMALITY	110	abnormality	35	24	4	44	2
7	ABORTION	126	abortion	0	60	22	41	2
8	ABSCESS	248	abscess	31	26	0	190	0
9	ABSORBED	279	absorbed	66	72	11	79	50
10	ABSORPTION	429	absorption	67	114	28	177	42
11	ABUSE	358	abuse	0	33	114	197	13
12	AC	282	ac	0	10	147	124	0
13	ACID	761	acid	140	177	15	334	94
14	ACIDOSIS	224	acidosis	9	30	3	167	14
15	ACIDS	301	acids	132	71	5	81	11
16	ACTIVATED	123	activated	19	24	11	66	2
17	ACUITY	154	acuity	0	33	47	73	0
18	ACUTE	2195	acute	173	431	151	1380	59
19	AD	386	ad	0	19	90	278	0
20	ADHERE	109	adhere	9	11	10	78	0
21	ADHERENCE	195	adherence	0	14	30	135	15
22	ADMINISTER	259	administer	1	45	53	149	10
23	ADMINISTERED	919	administered	5	134	44	700	35
24	ADMINISTERING	218	administering	0	17	44	152	4
25	ADOLESCENTS	184	adolescents	2	16	123	42	0

Figure 8: The Sample of Results Obtained using WLC 1.0 and WordSmith 5.0

Implication

The creation of field specific or focused word lists is believed to be put into good use by students when students actually take the initiatives to find the meanings of these specialised words in their first language. A word list is a presentation of L2 words which are tabulated side-by-side with the L1 translations, terms or definitions (Nation, 2001; Thornbury, 2002; Nakata, 2008). Learners are expected to create a personalised system for better vocabulary linking and retention (Godwin-Jones, 2010). Thus, it is up to the learners to determine how they wish to use an essential word list. In terms of words exposure, Thorbury (2002) suggests at least seven times of occurrence of a particular vocabulary item in order for the learner to learn it well, but Koprowski (2006) suggested at least 5 -16 times of words occurrence.

However, despite the disparate suggestions, learning or mastering a particular set of vocabulary depends strongly on the nature of word learning process which has always varied and been incremental (Huang & Liou, 2007; Nation, 1990). Any word list created serves only as an ultimate guide where teachers and learners should also teach and learn the words in various ways so that the words from the list can be beneficial to all. Despite various ways of creating word lists, the common criteria would be comparison of frequency, range and specialization and

perhaps representation in various or specific fields depending on the target corpus or word list intended to be created.

The word lists created from the software should also be learned with its collocations, some may call the lexical bundles. Educators should teach the essential words with its lexical bundles interactively to promote retention. Lexical bundles provide significant structural and functional pedagogical implications (Chen & Baker, 2010) and vocabulary retention can be enhanced by engaging learners with interactive database (Horst, Cobb & Nicolae, 2005). Thus, word list creation for specific field of study is deemed necessary for EAP and ESP context.

Conclusion

From the study, the new and innovative software is now available for all educators for various word lists creation purposes. Although there are many ways to generate a word list specifically for a certain discipline, better time economical software is desired as it increases the efficiency of a researcher and reduces human error in the study.

References

Anthony, L. (2009). *AntWordProfiler. [Computer software]*. http://www.antlab.sci.waseda.ac.jp/antwordprofiler_index.html.

Chung, T.M. & Nation, I.S.P. (2004). Identifying technical vocabulary. *System 32*, 2 251-263.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15 (2), 103-116.

Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL Word families in medical research articles (RAs). *English for Specific Purposes*, 26, 502–514. doi:10.1016/j.esp.2007.04.003.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2): 213-238.

Coxhead, A. & Hirsh, D. (2007). A pilot science word list for EAP. *Revue Française de Linguistique Appliqueé XII* (2), 65-78.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: Approaches to second-language vocabulary learning. *Language Learning & Technology, 14*(2), 4-11. Retrieved from http://llt.msu.edu/vol14num2/emerging.pdf

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. http://www.vuw.ac.nz/lals/staff/Paul_Nation

Heatley, A., & Nation, P. (1996). *Range [Computer software]*. Wellington, New Zealand: Victoria University of Wellington.

Hirsh, D., & Coxhead, A. (2009). Ten ways of focussing on science-specific vocabulary in EAP. *English Australia Journal, 25*(1), 5-16.

Horst, M., Cobb, T. & Nicolae, I. (2006). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology, 9*(2), 90-110. Retrieved from http://llt.msu.edu/vol9num2/horst/default.html

Huang, H. & Liou, H. (2001). Vocabulary learning in an automated graded program. *Language Learning & Technology*, *11*(3), 64-82. Retrieved from http://llt.msu.edu/vol11num3/huangliou/default.html

Hyland, K., & Tse, P. (2007). Is there an ''academic vocabulary"? *TESOL Quarterly*, 41(2), 235–253.

Kennedy, C. & Bolitho, R. (1984). *English for Specific Purposes*. London: Macmillan Publishers Ltd.

Koprowski, M. (2006). Ten Good Games for Recycling Vocabulary. *Internet TESL Journal*. Retrieved March 6, 2010, from http://iteslj.org/Techniques/Koprowski-RecylingVocabulary.html.

Li, Y., & Qian, D.D. (2010). Profiling the Academic Word List (AWL) in a financial corpus, *System*. doi:10.1016/j.system.2010.06.015

Martínez, I.A., Beck, S.C. & Panza, C.B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purpose,* 28, 183-198. doi:10.1016/j.esp.2009.04.003.

Menon, S. & Mukundan, J. (2010). Analysing Collocational Patterns of Semi-Technical Words in Science Textbooks. *Pertanika Journal of Social Sciences and Humanities, 18* (2), 241-258.

Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purpose, 25*, 235-256. doi:10.1016/j.esp.2005.05.002

Mukundan, J. (2009). *ESL textbook evaluation: A composite framework*. Köln, Germany: Lambert Academic Publishing.

Mukundan, J., and Anealka, A. (2009). Loading and distribution of the 2000 high frequency words in Malaysian English Language Textbooks for Form 1 to Form 5. *Pertanika Journal of Social Science and Humanities, 17*(2), 141-152.

Mukundan, J. & Anealka, A. (2007). A forensic study of vocabulary load and distribution in five textbooks within an entire Malaysian secondary school English language programme (Forms 1-5). *Pertanika Journal of Social Sciences and Humanities, 15* (2), 183-198.

Mukundan, J. and Menon, S. (2007). Lexical similarities and differences in the Mathematics, Science and English language textbooks. *K@TA (A Biannual Publication on the Study of Language and Literature), 9*(2): 91-111.

Mukundan, J., Roslim, N. (2009). Textbook Representation of Prepositions. *English Language Teaching*, 2 (4), 13-24.

Nakata, T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL Journal*, 20, 3-20.

Nation, P., (1990). *Teaching and Learning Vocabulary.* New York: Newbury House Publishers.

Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23 (1), 35-41.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge, Cambridge University Press.

Praninskas, J. (1972). *American University Word List*. London: Longman.

Reppen, R. (2001). Review of MonoConc PRO and Wordsmith Tools. *Language Learning & Technology , 5*(3), 32-36. Retrieved from http://llt.msu.edu/vol5num3/review4/default.html

Scott, M. (2010). *Version 5.0 WordSmith Tools*. Oxford: Oxford University Press.

Thornbury, S. (2002). *How to Teach Vocabulary*. Essex: Pearson Education Limited.

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 Specif. Purposes 28, 33-41. doi:10.1016/j.esp.2008.08.003

Wang, J., Liang, S.L., & Ge, G.C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27, 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 Purpose*, 28, 170-182. doi:10.1016/j.esp.2009.04.001

West, M. (1953). *A General Service List of English Words*. London: Longman.

Xue, Guoyi, & Nation, I. S. P. (l984). A university word list. *Language Learning and Communication*, 3(2), 2l5–229.