Advances in Social Science, Education and Humanities Research, volume 254
Eleventh Conference on Applied Linguistics (CONAPLIN 2018)

Comparing the Vocabulary Classification and Keyword Analysis Approaches to Determining Technical Vocabulary of Islamic Studies

Srifani Simbuka*
Tadris Bahasa Inggris
Institut Agama Islam Negeri Manado
Manado, Indonesia
*srifani.simbuka@iain-manado.ac.id

Wachyu Sundayana, Fuad A. Hamied
English Education Department
Universitas Pendidikan Indonesia
Bandung, Indonesia
swachyu@upi.edu, fuadah@upi.edu

Deny A. Kwary
English Department
Universitas Airlangga
Surabaya, Indonesia
kwary@yahoo.com

***Abstract*—Dealing with technical vocabulary is one of the essentials of English Language Teaching (ELT), yet it is still a major concern for learners and teachers who are bound to specific purposes of learning English. This paper reports on and compares two approaches to determining words to be included in a technical vocabulary list in the area of Islamic Religious studies (IRS): the vocabulary classification versus the keyword analysis. The first approach results in a list of four categories of words, namely high frequency words, academic words, technical vocabulary and low frequency words. Keyword analysis is the second approach, in which the end product is a list of words that saliently characterize certain texts. Drawing on the data of the Corpus of Islamic Religious Studies Textbooks in Indonesian Islamic State Institute, the two approaches are compared in terms of theoretical basis, method of analysis, and their end product. The target corpus (305,701 running words) is developed from core textbooks of Islamic religious studies whose content are aligned with the curriculum of Indonesian Islamic state universities. The merits and shortcomings of each approach are also elaborated. The result shows that a major difference between these two approaches is the presence or absence of some GSL word types in the two lists of technical vocabulary in Islamic religious studies.**

Keywords—corpus-based approach; Islamic religious studies; keyword analysis; technical vocabulary; vocabulary classification

I. INTRODUCTION

Pedagogy-oriented research on specially designed lists containing considerably high-frequency words in a corpus/collection of texts or wordlists [1], have mainly aimed to benefit the learning and teaching the vocabulary of second or foreign languages. The roles of wordlists in ESL/EFL are mainly in course design, materials development and testing [1-4]. Well-known wordlists of English general vocabulary are, among others, West's [5] General Service List (henceforward abbreviated as GSL) and the University Word List compiled by Xue and Nation in Nation [1].

Recent changes in the flowering borderless global education have paved the way for the increased needs of discipline-specific English vocabulary lists and ESL materials to facilitate ESL learning in academic specializations, which eventually facilitate the mastery of the core knowledge and or skills needed in those specializations [1]. Specialized vocabulary lists developed for finance [6], pharmacy [7,8], nursing [9,10], agriculture [11], newspaper [12], social sciences [13], plumbing [14] and engineering [15], are just a hand's full of the increasing number of technical vocabulary lists developed recently. These subject-specific wordlists are important addition to already existing academic wordlist such as Coxhead's Academic Word List (henceforward abbreviated as AWL) and the New Academic Word List [16].

This paper is aimed to compare two methods of creating technical vocabulary lists, namely technical vocabulary based on Nation's vocabulary classification and Scott's keyword analysis [1,17]. The comparative elaboration of these two quantitative/objective approaches to determining discipline-specific vocabulary lists are made in terms of the theoretical frameworks, methods of analysis and end products. This current study is a part of a larger PhD research project that used three layers of analysis in developing a technical vocabulary list in the discipline of Islamic Religious Studies (henceforward IRS), including the two approaches examined in this current paper. Such field-specific vocabulary list is important in assisting learners and teachers of English for Specific Purposes (ESP) to focus the learning to essential vocabulary of English related to their academic discipline [18,1].

II. VOCABULARY CLASSIFICATION

A. Theoretical Basis

The base-line of vocabulary classification proposed by Nation is the frequency of a word in any given corpus [3]. The first of category of this classification is obviously comprised of highly frequent words. In the most well-known list of general English compiled by *Michael West's General Service Lists (GSL)* [5], function words such as the article "the" is recorded at the top of the list. Other words ranked highest in GSL's 2000 most frequent English words list (GSL first 1000 and GSL second 1000 words) are the auxiliary "be", conjunctions such as "and" and preposition "to" [5].

The second category of Nation's vocabulary classification is Academic vocabulary or most frequent words in academic texts across various disciplines [3]. Academic word lists such as the University Word list compiled by Xue and Nation in Nation [1] and Coxhead's Academic List, enlisted words that are considered important for (mostly) non-native speakers of English who are pursuing studies in English-speaking countries [19]. Coxhead's list contained 570 word families, enlisting words that are not included in the GSL (first and second 1000) lists, but not too specific to certain specialization area. Some of the words in Coxhead's Academic were *analyze*, *concept*, *economic*, *identify*, etc. [19].

The next category of Nation's vocabulary classification is technical vocabulary [1,3]. Words that fall into this category are of two sub-categories: first, words specific to certain area but are known to common people outside its specific area for example commonly known medical vocabulary such as *influenza*, *muscle heart*, and *blood* [1]. The other sub-category is very specific vocabulary whose meanings signify that certain area of specialization. Words like *xiphoid*, and *arrhythmia* [1] belong to this sub-category of technical vocabulary. These words are more frequently in one specialized discipline rather than in other areas [3,1,12,20]. Other examples of technical words are those compiled in specialized dictionaries or encyclopedias compiled to meet the need of ESL/EFL learners studying in specific discipline in an English speaking universities or non-English speaking practitioners of a certain occupancies.

The later sub-category of specialized/technical vocabulary requires a considerably sufficient specialization knowledge to be understood and used in texts [20]. For the users of technical words, the mastery of the meaning of an item of this type of vocabulary does not only depend on their level of English, but also on their knowledge of the specific subject area in which the technical words/vocabulary are needed [3]. The knowledge of Linguistics, for example, is important for a student in this discipline to distinguish the difference between the word "morphology" in Linguistics textbooks and the one in a Biology article.

Low frequency words are the last category in Nation's classification [3]. As the name suggests, these are words that are very infrequently occur in a text. In relation to 'technical vocabulary', Nation stated that "One person's technical vocabulary is another person's low frequency word [3]". Some words, however, are very rarely occur in any text such as archaic words that belong to an older variation or a specific dialect of a language and even words that are infrequent due to some restriction based on social norms. Words from foreign languages can sometimes be included in this list. Proper nouns (names of people and places) are also classified as low frequency words, with an exception in specific genres such as fictional texts that certain proper names i.e. the names of the main characters of a novel may occur very frequently throughout the novel.

B. Method of Analysis

The methods of analysis of the two compared approaches to selecting technical vocabulary from a corpus are exemplified by drawing on the data of an ongoing-larger doctoral research in an Indonesian State Institute of Islamic Studies (henceforward SIIS) or Institut Agama Islam Negeri (henceforward IAIN). The target corpus containing 18,058 word types and 305,701 tokens was named the Corpus of Islamic Religious Studies Textbooks /CIRST [21].

The source texts for this specialized small-sized corpus were seven core textbooks in IRS. These textbooks were prescribed by the IAIN curriculum to be taught to all of its sophomore students as the university's obligatory basic IRS subjects. The IRS subjects were the sciences of the Qur'an/*ulum al-Qur'an*, the sciences of the hadith/*ulum al-hadith* and its methodologies, Islamic law and jurisprudence or *fiqh/ ushul fiqh*, Sufism/*tassawuf* and theology and philosophy/*Kalam*. All of the data source IRS textbooks were written in English by experts of IRS that include native-speakers of English and non-native speakers of English. The later writer group were presumably had considerably high competence in English since they were able to publish academic textbooks in English.

An initial processing of the data to be processed in any corpus-based approach study includes digitation of the target corpus source from which the analysis of word profiler-vocabulary classification is drawn from. The next step is converting the text into a plain text format using text conversion programs and cleaning-up unnecessary parts of the text such as page numbers and references. These steps are also crucial in other quantitative analysis of vocabulary classification, despite differences in the technique and corpus tools employed.

The unit of analysis in this present work is 'word type', which means the words type '*muslim'* and its plural inflection '*muslims'* are counted as two separate word types. The use of word type as the unit of counting in both of the two approaches to determining subject-specific vocabulary list follows Ward's argument that word type as unit of counting is suitable for wordlists developed to meet the needs of beginner learners [22]. Other studies used word families as the unit of analysis, for example Kwary and Artha [13] and Zhu among many others, based on the consideration that word families capture headwords and their inflectional and derivational forms, which were suitable for university level ESL/EFL students [12]. Yet, other studies mainly pioneered by Nation used families of lemma or 'flemma' in their wordlists [1].

Among the computer-based method to identifying technical vocabulary, Range see Kwary, was a pioneering popular computer program that was specially designed to filter out technical vocabulary from general high-frequency words and academic ones [6]. In fact, Vocabulary Classification [3] was the basis for the Range program to operate the analysis of word categories function [6]. Range can also be used to perform vocabulary load analysis of texts [14]. In this paper, however, the use of Range in determining technical vocabulary is not the focus of attention. Instead, the AntWordProfiler Program, a similar corpus computer program developed by Laurence Anthony is employed to demonstrate the identification of technical vocabulary in a specialized corpus elaborated in the next section [23].

1) Step 1: Classifying vocabulary using AntWordProfiler: The first technique to determining technical vocabulary used in this study is word profiling the target corpus using a corpus-computer software called the AntWordProfiler [23]. This step of profiling the types of vocabulary is a springboard to the actual step of selecting words that are qualified to be included in the (target) technical vocabulary list.

AntWordProfiler is a freeware for carrying out corpus linguistics analysis on profiling the content of corpus. Similar to the Range program, the word profile analysis was based on Nation's classification of words into four categories of vocabulary i.e. high frequency words, academic words, technical words and low frequency words [3]. By default, the software compared any target corpus against the GSL and AWL list and further categorized the words in the target corpus into three main groups: words that belong to the GSL, the AWL and "group not found" referring to words that do not belong to either of the previously mentioned groups.

This classification was done using the word profiler tool with a series of steps starting with running the AntWordProfiler, up-loading the texts data into the program, and analyze it using the word profiler tool. Next, the GSL (first 1000 word and the second 1000-word list) and AWL word list were up-loaded in the "Level list" section of the AntWordProfile platform, as the reference lists. These word lists served as the "filter" for sorting out the words of the data that belong to GSL and AWL.

Some operations represented in the dialogue boxes in the program were checked to generate the result, for example the 'statistics' box was checked to generate a list of the statistic of the word categories i.e. the "Level list' (the GSL and AWL) and potential IRV (words that were not listed in the level lists), and the 'Word Type' was checked for the result to contain all word types in the data, and so on. The data sorting was executed by clicking the tab "start". The result of this analysis was saved in txt files containing information about the files being processed, the words of the texts that belonged to the GSL first and second lists, the AWL list and a third list containing words that did not belong to those categories. The last 'group' of words also contained both potential technical words in Islamic studies/potential IRV and low frequency words. Statistical information of the word profile was also given that include the number and percentages of types and token of each group (GSL-first 1000 list, GSL second 1000 list, AWL list and "other" word list), as shown in figure 1 below:

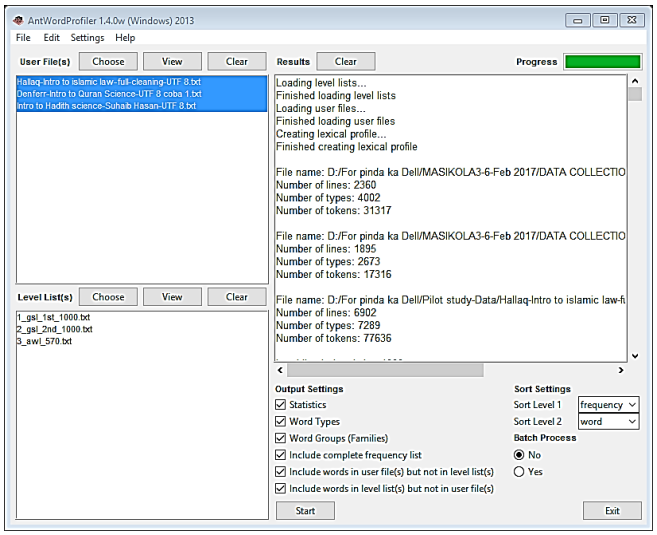

Fig. 1. Processing texts data source using AntWorProfiler-word profiler tool.

2) Step 2: Determining technical vocabulary: The "Groups not found base list" was then copied and pasted into an Excell file for manual selection by the researcher to select words that are most likely to be listed as potential technical vocabulary, which, in this case is exemplified using the data in the area of Islamic religious studies [21]. This selection was based on a set of criteria i.e. (i) the potential technical had to be non-member of GSL nor AWL; (ii) the words had to occur at least 9 times in the corpus (iii) through at least 3 of the 5 source texts topics in the target corpus.

C. End Product

The result of the first step in determining technical vocabulary based on the Vocabulary Classification [3] through the AntWordProfile-word profile tool showing that the target corpus was comprised of these categories [23]:

- High-frequency words (GSL) termed as "Level 1" and "Level 2" in the word profile result: 4,520 word types or 234,922 token.
- Academic vocabulary (AWL) termed as "Level 3" in the word profile: 1,877 word types or 20,091 token.
- Low frequency words and;
- Technical vocabulary was both enlisted in the "Level 0" or "group not found (in both GSL and AWL lists): 11,661 word types or 50,688 token.

The final product of the process was a list of technical vocabulary in IRS containing 262 word types/9,005 tokens (English and Anglicized Arabic words only), as seen in the table 1 below:

TABLE I. THE IRS TECHNICAL VOCAB ON THE BASIS OF VOCABULARY CLASSIFICATION

Rank	Word Types	Range	Frequency
1	Muslim	5	528
2	Prophet	5	401
3	Scholars	5	285
4	Muslims	5	253
5	Divine	5	136
6	one's	5	97
7	Guidance	5	74
8	Mosque	5	59
9	Prominent	5	57
10	Caliphs	5	49
11	Fasting	5	43
12	Scholar	5	42
13	well-known	5	42
14	Reference	5	40
15	Entitled	5	38
16	prophet's	5	37
17	Prophets	5	33
18	Believers	5	29
19	Genuine	5	29
20	Authentic	5	25
262	Unbelievers	3	9
The complete list is attached in the appendix 1			

III. KEYWORD ANALYSIS

A. *Theoretical Basis*

Another way of selecting technical vocabulary is by finding out whether or not a word' is "key" in a given text(s). For a word to be attributed as a "keyword" of a text, it has to occur with an unusually high frequency (positive keyword) or unusually low frequency (negative keyword) relative to its occurrence in a reference corpus [24,25,17]. The reliance on the statistical test to determine a word's "keyness" against a reference corpus challenged a more conventional method for determining keywords based on subjective judgment on the basis of "cultural importance" of certain words in certain text. Keywords of a particular text is, according to Scott and Tribble of three types [17]: proper nouns, words that indicate the "aboutness" of a text and high frequency words indicating style. The comparison between the two corpora, by Scott's typology of keywords, should then unpack the most significant lexical differences between them, with regard to "aboutness" and style [26].

B. *Method of Analysis*

Similar to vocabulary classification, Keywords analysis is also supported by a specially designed computer program. Scott developed WordSmith to extract keywords automatically with the program's keyword tool [27]. The statistical test used in this non-freely accessed software is log-likelihood or chi-squared tests to compare two word lists in order to generate keywords [28,13,6,25]

Instead of using WordSmith, the keyword analysis exemplified in this paper is done with another computer program called AntConc [29]. This freeware has been utilized in a number of studies of developing keyword list functioned as technical word list such as the most recent one conducted by Weisser [30]. Using the texts and the procedure in the CIRST [21], the keyword list of Islamic Religious studies (IRS) was produced following the protocol as listed below:

- There was no "Stop List" was used in keyword tool which means that unlike the analysis using AntWordProfiler, the input data of 18,875 Types were not filtered by any stop list. Antconc also reported a slightly different number of word/Types than AntwordProfiler software.
- The Antconc keyword tool required a reference corpus to which the target corpus (CIST-IISI) was compared, in order to generate a keyword list. The reference corpus in this particular analysis was the British Academic Written English (BAWE).
- The 'keyness' value was calculated by the Antconc software using Pearson chi-squared or a log likehood statistical tests, which were both available on the keyword tool selection of the Antconc software. In this particular study, a chi-square test was used to compare the distribution by type of corpus i.e. target (CIST) versus reference (BAWE) corpus of the 18875 word types in the Target corpus (CIST-IISI)- example chi square to see the "difference" of the frequency of the word type "prophet" in CIST with its expected occurrences /frequencies in CIST and BAWE.

The procedure of running the AntConc keyword program was, to a certain extent, simple. The texts files (data source) were loaded into Antconc and analyzed using the "Keyword tool". In the "tool setting" dialog box, BAWE was specified as the reference corpus. The "Token definition" dialog box was set to accommodate dash and apostrophe to be considered as part of a "word". This step was as seen figure 2 below:

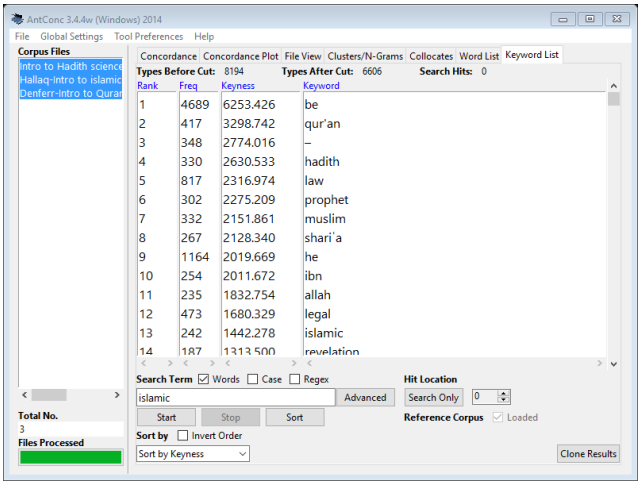

Fig. 2. Processing texts data source in the AntConc keyword tool.

C. *End Product*

The result of AntConc keyword tool that compares the frequency or words between a target and a reference corpus resulted in a list showing a descending rank of keywords of the target (in this case, the CIRST). The 'strongest' or most statistically significant keywords appear first in the list.

Presented in Picture 3 below was a list of the result from the Antconc Keyword analysis that ranked the words/Types of the CIST on the basis of the "keyness" value [21].

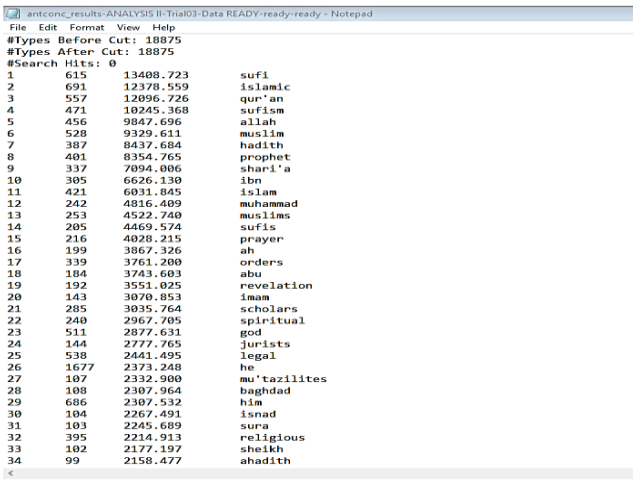

Fig. 3. The result from the Antconc Keyword analysis.

There was no "Stop list" used in the word list tool setting. The resulting wordlist therefore contained words belonging to GSL and AWL. The result list was therefore had to be sorted out manually, in order to identify GSL and AWL. In this manual process, non-English words were deleted from the original list above (figure 3), therefore the Arabic word "*sufi*" that initially appeared as the strongest keyword of the target corpus (CIRST) did not appear in the final result of the keywords of IRS list, showed table 2 below:

TABLE II. THE IRS TECHNICAL VOCABULARY ON THE BASIS OF KEYNESS

Rank	Word Types	Keyness	Frequency
1	Islamic	**12,378.6**	691
2	Sufism	**10,245.4**	471
3	Muslim	**9,329.6**	528
4	Prophet	**8,354.8**	401
5	Islam	**6,031.8**	421
6	Muslims	**4,522.7**	253
7	Sufis	**4,469.6**	205
8	Prayer	**4,028.2**	216
9	Revelation	**3,551.0**	192
10	Scholars	**3,035.8**	285
11	Spiritual	**2,967.7**	240
12	God	**2,877.6**	511
13	Jurists	**2,777.8**	144
14	Narrated	**2,441.5**	538
15	mu'tazilites	**2,332.9**	107
16	Religious	**2,214.9**	395
17	Companions	**2,124.3**	111
18	Arabic	**2,051.6**	111
19	Prayers	**2,046.7**	114
20	Law	**1,732.2**	787
21	Reporters	**1,719.9**	85
22	Bless	**1,698.2**	84
23	Obligatory	**1,625.3**	94
24	Bequest	**1,569.8**	72
25	Theology	**1,564.9**	132
The complete table was affixed in Appendix.2			

IV. DISCUSSION

The keywords above contained different word types that were absent in the technical vocabulary list created on the basis of vocabulary classification. By looking at the top 20 word types of the Keywords list and the Technical vocabulary list of IRS, it is revealed that there were word types in the IRS Keywords List that did not appear as high ranked word types in the other list. These word types were *Islamic*, *sufism*, *Islam*, *sufis*, *prayer*, *revelation*, *spiritual*, *God*, *jurists*, *narrated*, *mu'tazilites*, *religious*, *companions*, *Arabic*, *prayers* and *law*. Meanwhile, the word types *divine*, *one's guidance*, *mosque*, *prominent*, *caliphs*, *fasting*, *scholar*, *well-known*, *reference*, *entitled*, *prophet's*, *prophets*, *believers*, and *genuine* were enlisted at the top 20 of highly frequent and evenly distributed word types in the IRS Technical Vocabulary List. These word types, however, were absent at the IRS Keyword List top 20 word types with the highest "keyness" value. Of the top 20 word types exemplified from both of the lists of IRS's most important word types, there were only four (*muslim*, *prophet*, *scholars*, and *muslims*) that were shared by both of them, although with a slight difference in their ranking position in each of the analyzed word lists.

Another point of difference in the end products of these two approaches of determining technical vocabulary was that the IRS Technical Vocabulary List were free from high-frequency words (GSL) and academic vocabulary (AWL). As the result of using GSL and AWL as the "stop list", the word types of IRS Technical Vocabulary Lists were automatically "filtered out" from the two previous categories of Nation's classification of vocabulary [3].

On the other hand, high-frequency and academic words, as well as some (generally) low-frequency words, proper nouns and function words were all included in the end product of keyword analysis, i.e. the IRS Keyword List. This was due to the fact that keyword analysis did not compare the content of the target corpus with existing "word lists" (part of a corpus), but comparing the whole target corpus to another whole corpus. The presence of high-frequency and academic words in the IR Keyword List confirmed claim that "technical vocabulary" might include words that had both general and specialized words [1,31].

Based on these findings, therefore, it is concluded that the IRS Technical Vocabulary on the basis of Vocabulary Classification and the IRS Keyword List were "slightly different" with regard of their end products.

V. CONCLUSION

The motivation behind corpus-linguists' effort in compiling word lists, would they be general vocabulary or discipline-specific vocabulary are mostly to weigh the lists' pedagogical benefit the learning of second or foreign language vocabularies. Specific to the endeavors of creating technical vocabulary lists, various subjective, objectives even combined methods were employed and tried out with the best intention of producing technical vocabulary lists that are 'accurate' and beneficial for learners. This paper, thus compared two quantitative (objective) approaches to determining discipline-specific

vocabulary lists from two relatively 'different' stand points i.e. vocabulary classification [3] versus keyword analysis [17].

An underlying similarity between vocabulary classification and keyword analysis to determining technical vocabulary was on their emphasis on frequency measure. The salience on frequency was made by referring to the roles of frequency measure in determining pedagogical benefit of word list on (ESL/EFL) vocabulary learning. The difference at using and interpreting frequency measure as the tool of analysis of each of these two approaches, however, resulted quite expectedly in 'different' content of the discipline-specific vocabulary lists they produced. Despite their contrasting nature of methodology and (consequently) final products, both of these lists were of invaluable sources for ESL/EFL learners for acquiring English vocabulary in the area of Islamic (religious) Studies. Therefore, it is recommended that further research will extent this work and reveal ways to negotiate between the differences of these two approaches, and or with other quantitative as well as qualitative approaches to produce even higher validity and accuracy of technical vocabulary lists.

REFERENCES

[1] I.S.P. Nation, Making and Using Word Lists for Language Learning and Testing. (I. S. P. Nation, Ed.) (e-Book). John Benjamins, 2016.

[2] L. Flowerdew, Corpora and Language Education. Palgrave Macmillan, 2012

[3] I.S.P. Nation, Learning Vocabulary in another Language (First). Cambridge: Cambridge University Press, 2001.

[4] I.S.P. Nation, "Designing reading tasks to maximise vocabulary learning," Applied Research on English Language, vol. 3, no. 5, pp. 1–8, 2013.

[5] M. West, GSL-AWL word list.pdf. In A general service list of English. London: Longman, Green & Co, 1953.

[6] D.A. Kwary, A hybrid method for determining technical vocabulary. System, vol. 39, no. 2, pp. 175–185, 2011.

[7] L. Grabowski, "Register Variation Across English Pharmaceutical Texts: A Corpus-driven Study of Keywords, Lexical Bundles and Phrase Frames in Patient Information Leaflets and Summaries of Product Characteristics," Procedia - Social and Behavioral Sciences, vol. 95, pp. 391–401, 2013

[8] L. Grabowski, "Keywords and lexical bundles within english pharmaceutical discourse: A corpus-driven description," English for Specific Purposes, vol. 38, pp. 23–33, 2015.

[9] A.F. Mohamad and Y.J. Ng, "Corpus-based studies on nursing textbooks," Advances in Language and Literacy Studies, vol. 4, no. 2, pp. 21–28, 2013.

[10] M.N. Yang, "A nursing academic word list," English for Specific Purposes, vol. 37, no. 1, pp. 27–38, 2015.

[11] V.L. Muñoz, The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. English for Specific Purposes, vol. 39, pp. 26–44, 2015.

[12] J. Zhu, The Technical Vocabulary Of Newspapers. The University of Western Ontario, 2017.

[13] D.A. Kwary and A.F. Artha, "The Academic Article Word List for Social Sciences 1," Mextesol Journal, vol. 41, no. 4, pp. 1–11, 2017.

[14] A. Coxhead and M. Demecheleer, "English for Speci fi c Purposes Investigating the technical vocabulary of Plumbing," English for Specific Purposes, vol. 51, pp. 84–97, 2018.

[15] N.Y. Jin, L.Y. Ling, C.S. Tong, N. Sahiddan and A. Philip, Development of the Engineering Technology Word List for Vocational Schools in Malaysia. International Education Research, vol. 1, no. 2001, pp. 43–59, 2013.

[16] D.E.E. Gardner and M. Davies, A New Academic Vocabulary List, (August 2013), pp. 305–327, 2014.

[17] M. Scott and C. Tribble, Textual Patterns: Key words and corpus analysis in language education (1st ed.). Philadelphia: John Benjamins, 2006.

[18] M. Abudukeremu, A Corpus-Based Lexical Study of the Frequency, Coverage and Distribution of Academic Vocabulary in Islamic Academic Research Articles. International Islamic University Malaysia, 2010.

[19] A. Coxhead, "A New Academic Word List," TESOL Quarterly, vol. 34, no. 2, pp. 213–238.

[20] T.M. Chung and P. Nation, "Identifying technical vocabulary," System, vol. 32, no. 2, pp. 251–263, 2004.

[21] S. Simbuka, F.A. Hamied, W. Sundayana, and D.A. Kwary, "A Corpus-based study on the technical vocabulary of Islamic religious studies," pp. 1–15, (n.d.).

[22] J. Ward, A basic engineering English word list for less proficient foundation engineering undergraduates. English for Specific Purposes, vol. 28, no. 3, pp. 170–182, 2009.

[23] L. Anthony, AntWordProfiler. Tokyo, Japan: Waseda University, 2014b, [Online]. Retrieved from: http://www.antlab.sci.waseda.ac.jp/.

[24] P. Baker, A. Hardie, and T. Mcenery, A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006a.

[25] A. O'Keeffe and M. McCarthy, The Routledge Handbook of Corpus Linguistics. (A. O'Keeffe & M. McCharty, Eds.) (1st ed.). New York: Routledge, 2010.

[26] P. Baker, "Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis," Journal of English Linguistics, vol. 32, no. 4, pp. 346–359, 2004.

[27] M. Baker Scott, WordSmith Tools version 7 , Stroud: Lexical Analysis Software. 2017. [Online]. Retrieved from: https://lexically.net/publications/citing_wordsmith.htm.

[28] P. Baker, A. Hardie, and T. Mcenery, A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006b.

[29] L. Anthony, AntConc. Tokyo, Japan: Waseda University. 2014a [Online]. Retrieved from: http://www.antlab.sci.waseda.ac.jp/.

[30] M. Weisser, Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis (1st ed.). Malden: Wiley & Blackwell, 2016.

[31] C. Sutarsyah, P. Nation, and G. Kennedy, "How Useful is EAP Vocabulary for ESP? A Corpus BAsed Case Study" RELC Journal, vol. 25, no. 2, pp. 34–50, 1994.