Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, 2020, 9(3) (Previously Published under the Title: Maritime English Journal)

Developing an Academic Word List for the Students of Health Information Management: A Corpus Study

[1] Danial Shirzadi
[2] Hamid Reza Dolatabadi*

IJEAP- IJEAP-2003-1517
Received: 2020-03-25 Accepted: 2020-09-11 Published: 2020-09-14

Abstract

Importance of discipline specific vocabulary knowledge is well perceived. Despite the importance of the issue, there is a dearth of empirical research to unravel frequent academic vocabulary in the field of Health Information Management. To fill this research gap, the present study drew on a corpus of research articles and course books written in this field to underscore the concept of Academic Word List. To do so, a corpus of 2,264,981 running words analyzed by Range software package. In the process of research, we proposed a new discipline specific word list specially tailored for the students of health information management which can cover for 15.09% of all tokens in the corpus. This proportion of coverage is an improvement over previous academic word lists. Accordingly, it is hoped that the findings of the present study could contribute to students, teachers, material developers and researchers in this field of study.

Keywords: Research articles, Course Books, Discipline, Vocabulary Knowledge, Tokens

1. Introduction

Importance and difficulty of vocabulary intake and usage is well documented in the field of second/foreign language learning (Cobb & Horst, 2002; Hirsh & Nation, 1992; Nagy &Townsend, 2012; Townsend & Kiernan, 2015). Shaw (1991) related most of students' problems and difficulties in writing and reading to their limited vocabulary knowledge and the fact that students, themselves, want to study vocabulary more than other realms of language (Hsu, 2011; Leki & Carson, 1994).

Nation (2001), in his pioneering work, divided vocabularies in an academic field to four categories: high-frequency words, academic words, technical words and low-frequency words. High-frequency words allude to those essential English words which constitute the lion's share of informal discussion or discourse and additionally all the running words in a wide range of composing. Technical words are the ones utilized as a part of a specific field, which are extensively not quite the same from one field of study to the other. Low-frequency words are scarce utilized terms. The function of academic word lists, somewhere between the high-frequency and technical words, is of great importance in every educational context. Thus, getting these words covered is by all accounts basic when students are trying to study texts of special fields. Academic vocabulary has been defined by Farrell (1990) as:

Formal, context-independent words with a high frequency and/or a wide range of occurrence across scientific disciplines, not usually found in basic general English. Academic words, not usually found in basic general English texts, refer to words that account for a relatively high proportion of running words in all academic texts, courses, words with high- frequency across scientific disciplines (Farrell, 1990, p. 11).

Coxhead (2000) believes that academic vocabulary can cause a great deal of difficulty because of their obscure meaning to the learners of a special discipline while this is usually not the case with technical vocabulary (i.e. Learners are familiar with most of the technical words in their field of study).It is believed that, academic vocabulary (the most important of which is Coxhead's

[1] PhD candidate of TEFL, Shirzadidanial@yahoo.com; Department of English Language and Literature, Faculty of Humanities, Arak University, Iran.

[2] Assistant Professor in TEFL (Corresponding Author), H-dowlatabadi@araku.ac.ir, Department of English Language and Literature, Faculty of Humanities, Arak University, Iran.

academic word list containing 570 word families) accompanied with general vocabulary (West's general service list containing 2000 word families) account for almost 85% of every academic text (Coxhead &Nation, 2001; Crompton, 2013;Young, 2015; Ward, 1999). In addition, more recent studies revealed that for greater precision in understanding academic texts (98%) learners need to know approximately 8000 to 9000 word families (Laufer & Ravenhorst- Kalovski, 2010; Nation, 2006). Seemingly, importance of academic words is no secret but the questions of how many and which have remained unanswered.

Early attempts to reach a scientific list of vocabulary for teaching were made with the advent of corpus linguistics by West's (1953) General Service List (GSL). In Nation's (2001) study, the results indicated that the list of 2000 high frequency words covers for the 80% of vocabularies in all texts (Nation, 2001). Accordingly, this list is of great importance and it can be mentioned as the starting point to scientific teaching and studying of vocabulary.

With maturity of English for Academic Purposes and English for Specific Purposes fields of study, the dearth of an academic world lists (general and discipline specified) was perceived. One of the earliest attempts was made by Coxhead (2000). She analyzed a corpus of 3.5 million words coming from four different academic fields: The Arts, Commerce, Law, and Natural Science. She set three criteria in her corpus study:

1. "frequency: occurrence of 100times through the entire corpus
2. range: occurrence of at least 10 times in each of the sub disciplines
3. specialized occurrence: to be out of West's general service list of high frequency vocabulary" (p. 44)

Finally, she came up with a list of 570-word families which has been divided to 10 sub-lists. Each sub-list consists of 60-word families except for the 10^{th} which contains only 30. Coxhead (2000) claimed that her list would account for 10% of all academic texts. While its coverage was versatile among different four sub-disciplines (i.e. it was 9.3%, 12%, 9.4% and 9.1 % for Arts, Commerce, Law and Science, respectively). She argued that Academic Word List (AWL) is of crucial importance for students of different academic fields as it can cover for almost 90% of running words in academic texts when you combine it with West's GLS. Positive effects of AWL have been documented and confirmed thoroughly by some researchers (Huntley, 2005; Li & Qian, 2010).

Despite all the success which AWL brought in academic vocabulary acquisition, it was not without critiques and flaws. Excluding medical texts (Chen & Ge, 2007), versatility of meaning and coverage of the words across different sub-corpora (Hyland &Tse, 2007) can be mentioned as some of those shortcomings.

After Coxhead's (2000) pioneering work, many researchers from different fields of studies tried to examine the coverage rate of AWL in specialized fields of study and turned the "focus on the academic vocabulary closely related to disciplines" (Liu &Han, 2015, p. 2) to come up with Field-specific academic word lists (Martínez, Beck, & Panza, 2009) which can be more related to specialized fields of study or even have a higher coverage percentage.

2. Review of the Literature

Coxhead (2011), in an insightful article, tried to justify the necessity and usefulness of an academic word list and mentioned some seminal works on the conformity of AWL to different corpora. The following chart can be a cogent and concise summary of researches which have been done on Awl's coverage over different disciplines.

Table 1: Studies Investigating AWL Distribution in Texts (adapted from Coxhead, 2011, p. 357)

Study	Corpus	Number of running words	Percent coverage of the AWL
Coxhead (2000a, 2000b)	Fiction	3.5 million	1.4
Coxhead (unreported)	Newspapers	1 million	4.5
Cobb & Horst (2004)	Learned section of the Brown corpus (Francis & Kucera, 1979)	14, 283	11.60
Hyland & Tse (2007)	Sciences, engineering, and social sciences, written by professional and student writers	3,292,600	10.6
Chen & Ge (2007)	Medical research articles	190,425	10.073
Konstantakis (2007)	Business	1 million	11.51
Coxhead & Hirsh (2007)	Science	1.5 million	8.96
Ward (2009)	Engineering	271,000	11.3
Martı´nez, Beck, &Panza (2009)	Agricultural sciences research articles	826,416	9.06
Vongpumivitch, Huang, & Chang (2009)	Applied linguistics research papers	1.5 million	11.17
Li & Qian (2010)	Finance	6.3 million	10.46
Coxhead, Stevens, &Tinkle (2010)	Pathway series of secondary science textbooks	279,733	7.05

Researchers pursued studying AWL's conformity and coverage in different sub-disciplines and documented more specific and field dependent academic vocabulary lists up to date. For instance, Moiniand Islamizadeh (2016), following Vongpumivitch, Huang, and Chang (2009), investigated AWL in a 4-million-word corpus of applied linguistics articles. They suggested that AWL accounted for 10.18% of the words in applied linguistic research articles corpus which was lower than Coxhead's (2011) study (11.17%). They also devised a list of 224 frequent word families out of AWL and GLS which accounted for 18.51% of words in the aforementioned corpus.

Lei and Liu (2016) drew on Gardner and Davies's (2014) method and criticized Coxhead (2011) for not including high frequency general words, developed a new medical academic word list. They suggested that their list is shorter up to 53% than existing lists devised before (Chen & Ge, 2007; Wang, Liang, &Ge, 2008) and got an even better coverage. Combining a " 2.7 million-word corpus of medical academic English" and "a3.5 million-word corpus of medical English textbooks", they studied word families' frequency based on an eclectic framework consisting of Coxhead and Gardner and Davies's analysis premises: Minimum frequency, Frequency ratio, Range ratio, Dispersion, Discipline measure, Special meaning criterion for general high-frequency words. Finally, they highly recommended their new criterion framework for analyzing corpora and the importance of high frequency general words as they "often have special meanings in the discipline" (Lei & Liu, 2016, p. 49).

Todd (2017) emphasized the importance of vocabulary teaching in ESP classes but hesitated about the value of current word lists. He also paid attention to syntactic meanings of words which are field specific and different from general meaning of that word. He believed that "the main criterion for choosing which words and meanings should be included on the final list is opacity. This criterion should identify those words for which the learners would gain the greatest benefit from a teacher's help" (Todd, 2017, p. 38) and he called them "The opaque word". Based on this conceptual framework, he analyzed a 1.15-million-word corpus of engineering text books and came up with a list of 186 opaque items.

Liu and Han (2015) questioned the usefulness of AWL in the field of environmental science and they proved that their specialized academic word list significantly outperformed AWL in covering words used in environmental science research papers by 3.09%. They borrowed the word

"usage" from Juilland and Chang-Rodríguez, (1964) and introduced the concept of "optimized usage" as the featured criteria of vocabulary inclusion. Method of calculating optimized usage was defined as: "First, we removed a word's highest frequency value in the ten subject areas and then we calculated the word's usage, which is called 'optimized usage' in the study. (Liu & Han, 2015, p. 7).

As it has been shown briefly by preceding paragraphs after Coxhead's (2000) groundbreaking work, importance and necessity of developing field specific academic vocabulary has been felt more and more (Hyland, 2002; Martinez et al., 2009; Paquot, 2007; Samraj, 2002; Ward, 2009). Since then, so many disciplines and academic fields of study have been searched for high frequency vocabulary which has been called academic vocabulary before to empower students of different fields of academy in order to read academic writings. An academic word list designed especially for the students of *Health Information Management*, which has been introduced as a well-established field of academy, can be a great help to the students of this field who are supposed to read articles in English to complete their courses or write their thesis. Despite the importance of the issue, there is a dearth of empirical study to unravel frequent academic vocabulary in the field of *Health Information Management.* To fill this research gap, the present study drew on a corpus of research articles and course books written in this field to shed some lights on the concept of AWL and discipline specific academic words. In order to do that, the following research questions were answered:

Research Question One: To what extend AWL and GSL word families are used in (covered for) *health information management* corpus (HIMC)?

Research Question Two: What are the most frequent AWL and GSL word families used in *health information management* corpus (HIMC)?

Research Question Three: What are the most frequently occurring word families in the corpus of *health information management* that are not listed in AWL and GSL?

Research Question Four: Compared to AWL, does the new Health Information Management Academic Word List (HIMAWL) have a better coverage of *health information management* corpus (HIMC)?

3. Methodology

3.1. The Corpus

In order to develop Health Information Management Research Article Corpus (HIMRAC), two content experts (PhD holders and university professors) were consulted and they came up with a list of15 journals. Pursuing the investigation for compiling a more representative corpus, content experts reduced the list to 5 most accredited journals which have been published for more than 10 years and have been hosted by international publishers like Elsevier, Sage, Tailor and Francis, Springer and Pub Med. They also enjoyed an impact factor above 1.00. Representativeness, specificity of corpus, use of whole documents, and availability in electronic form were also among the selection criteria (Barnbrook, 1996; Sinclair, 1991, 2005). All articles were published between the time span of 2000 to 2017 and they should have a balanced length of 2000 to 7000 words. Applying the abovementioned limitations, we came up with 250 research articles. Table 2 concisely demonstrates the number of journals, research articles and running words.

Table 2: List of Related Journals

Name of the journal	Number of articles	Number of words
International Journal of Medical Informatics	50	317,482
BMC Medical Informatics and Decision Making	50	249,186
Health Information Management Journal	50	358,383
Informatics for Health And Social Care	50	286,742
perspectives in health info management	50	161,491
Total	250	1,373,284

In order to develop Health Information Management Course Books Corpus (HIMCBC), we consulted with two field experts and came up with 5 important course books which are currently being taught worldwide. To ensure the comprehensibility of our corpus, the researcher surfed some colleges 'syllabi (e.g. Midland College: available at www.midland.edu/docs/public_information/paci/hb2504/syllabi/.../HITT1311.pdf). These course books were downloaded and prepared to be analyzed similar to HIM research papers corpus. Table 3 demonstrates the name of books and their running words.

Table 3: List of Related Text Books

Name of the book	Author	Number of words
Implementing an Electronic Health Record System	James M.Walker Eric J. Bieber Frank Richards	80,667
Essentials of Health Information Management: Principles and Practices	Michelle A. Green Mary Jo Bowie,	157, 143
Case Studies in Health Information Management	Charlotte McCuen, Nanette B. Sayles, Patricia Schnering,	78,873
Electronic Health Records	Byron R. Hamilton	117,572
Health Information Management Technology an Applied Approach	Nanette B. Sayles	402,927
Total		837,182

3.2. Data Collection

Having downloaded the research articles in PDF format, the researcher copied all mentioned sections of every research article (abstract, introduction, methods, results and discussion) into a word file and then converted it to a text file so it can be read by Range software package (downloadable at *http://www.vuw.ac.nz/lals/staff/Paul_Nation*.). Range software, developed by Heatley, Nation, and Coxhead (2002), has been used widely to study words frequency and range in different corpora. It has the ability to sort out word families and compare them with GSL and AWL. Aside from mentioned sections of research articles, all other sections like tables, footnotes, acknowledgements, conclusions, bios, references, and appendixes were eliminated in order to standardize the corpus.

In order to achieve a more representative source the two corpuses were integrated and the final draft of health information management corpus (HIMC) was devised. The final corpus encompasses 2,210,466 running words coming from the above-mentioned research articles and course books.

3.3. Data Analysis

Following Coxhead (2000), this study considers three criteria for inclusion of words of health information management academic word list (HIMAWL): frequency, range and specialized occurrence. As far as frequency criterion is concerned, Coxhead selected word families with more than 100 frequencies along her 3.5-million-word corpus. Accordingly, in HIM only words with the frequency of 66or more were selected. Regarding range criterion, only word families which occurred in at least half of the journals and books were selected (i.e. the range factor was set at 5). Though, the major controversial issue in many studies was the concept of specialized occurrence. Some researchers ignored the distinctions between general and specialized usage of words (Billuroglu & Neufeld, 2005; Valipouri & Nassaji, 2013) but Lei & Liu criticized this approach and believed "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). In present study, the latter approach was followed to compile the first word list which is the list of most frequent words used in GSL and AWL. But the former approach was considered in compiling the discipline specific academic word list (HIMAWL) for the students of HIM (which is a combination of most frequent words of AWL and words which were not in any lists but met the three aforementioned criteria). To put it in nutshell, word selection criteria of the present study are frequency and range for creating the first word list but for the second list we go through all three criteria of frequency, range and specialized occurrence.

The present study only included content words. Functional words and abbreviations were eliminated. Word families defined as the root word plus its inflections and derivations by Bauer and Nation (1993) were considered as target units. Frequency, range and distribution of word families in research articles and course books corpora of HIM field of study were quantitatively calculated using Range software. Accordingly, the two word lists (AWL + GSL, and HIMAWL) were prepared and compared to the other word lists.

4. Results

This study aimed to devise an academic word list for the students of health information management. Towards this aim, a corpus of academic textbooks and research papers were compiled and analyzed. Results are illustrated in the following table.

Table 4: Coverage of Lexical Items in the HIMC

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	1384207/62.62	3309/ 5.85	981
Two	165202/ 7.47	1996/ 3.53	816
Three	298477/13.50	2557/ 4.52	570
Not In The Lists	362580/16.40	48706/86.10	?????
Total	2210466	56568	2367

As Table 4 shows, the corpus of the present study consists of 2,210,466 tokens, 56568 types, and 2367word families. The first and second GSL word lists accounted for 70.36 percent of running words in HIMC. Coxhead's (2000) AWL added this coverage by 13.50 percent and lived it up to 88.86; while 16.40 percent of running words were not in any lists.

It should be noted here that One- and Two-word lists belonged to GSL, and the others were related to AWL. The high coverage of GSL and AWL in the present study indicates that these word families are passim in HIMC and they are of great importance for the students of HIM. The following table is to summarize the coverage of these two-word lists over different corpora.

Table 5: Coverage of GSL and AWL in Other Studies

Word lists	The present study	Coxhead (2000)	Martinez et al. (2009)	Li and Qian (2010)	Khani and Tazik (2013)	Valipouri and Nassaji (2013)	Liu And Han (2015)
GSL	70.36	76.1	67.53	72.63	76.4	65.46	70.61
AWL	13.50	10	9.06	10.46	11.96	9.96	12.82
GSL+AWL	88.86	86.1	76.59	83.09	88.00	75.42	83.43

As it is shown in table 5, the coverage of the two mentioned word lists on HIMC was more than that of in other studies. Thus, GSL and AWL were considered as prerequisite to read and comprehend texts (Coxhead, 2000; Ward, 2009).

In addition, the current study was set to investigate the most frequent GLS and AWL word families used in *health information management* corpus (HIMC). In order to devise a list of most

frequent GSL and AWL word families each of them had to pass through some criteria. First of all, they should occur in at least 5 corpora or more; after that, they had to occur 66 times or more through the whole corpus. The HIM corpus was analyzed with the range software package and it provided us with a long list of vocabularies in which we eliminated those words which did not meet the mentioned criteria. Using predetermined word selection criteria, the present study came up with a list of words encompasses 1006 items. The most frequent word in the corpus was *use* which used 19049 times in all 10 journals and textbooks corpora while the least frequent one was *poison* which occurred 68 times in 5 sub-corpora. The most 30 frequent words in HIMC which coincided with GSL and AWL are listed below.

Table 6: The First 30 Most Frequent Types in The Present Corpus.

rank	Types	range	frequency	%	rank	types	range	frequency	%
1	USE	10	19049	0.86	16	CODE	10	4689	0.21
2	HEALTH	10	17164	0.77	17	REPORT	10	4532	0.20
3	INFORM	10	17111	0.77	18	ORGANIZE	10	4247	0.19
4	DATA	10	13428	0.60	19	PROCESS	10	4240	0.19
5	SYSTEM	10	12287	0.55	20	DOCUMENT	10	4204	0.19
6	PATIENT	10	11431	0.51	21	NEED	10	4084	0.18
7	RECORD	10	10080	045	22	RESULT	10	4073	0.18
8	CARE	10	9496	0.42	23	DEVELOP	10	3921	0.17
9	PROVIDE	10	8339	0.37	24	RESEARCH	10	3813	0.17
10	MEDICAL	10	7259	0.32	25	ACCESS	10	3797	0.17
11	MANAGE	10	7196	0.32	26	REQUIRE	10	3755	0.16
12	STUDY	10	6709	0.30	27	QUALITY	10	3723	0.16
13	HOSPITAL	10	5788	0.26	28	SUPPORT	10	3580	0.16
14	INCLUDE	10	5490	0.24	29	PRACTISE	10	3448	0.15
15	SERVICE	10	5378	0.24	30	IDENTIFY	10	3408	0.15

***Bold typed words** occur in Coxhead's (2000) AWL

These words enjoyed a wide range across all journals and books (i.e. all of them have a range of 10) and they also occurred 215,722 times through the corpus which made up 9.75% of the corpus. Interestingly, 8 word types out of 30 most frequent words were listed in AWL. Out of 1006 items of HIMC word list, 675 types occurred in the entire sub corpora. The total frequency of these words is 692,117 which accounted for 31% of the whole corpus.

Moreover, the present study explored the frequently occurring word families in the corpus of *health information management* are not listed in AWL and GSL. To this end, academic word families extracted from GSL and AWL were analyzed using range software. The results indicated that 429 word families of the mentioned word list were coincided with AWL which accounted for 42.64 % of the list. This percentage was 39.66 and 17.69 for the first and second part of GSL respectively. The results are illustrated in the following table.

Table 7: Proportion of Word Families from GSL And AWL Coincided with the New World List.

WORD LIST	TOKENS/ %		TYPES/ %		FAMILIES
One	399/	39.66	399/	39.66	399
Two	178/	17.69	178/	17.69	178
Three	429/	42.64	429/	42.64	429
Not In The Lists	0/	0.00	0/	0.00	?????
Total	1006		1006		1006

Table 4 indicated that out of 2,210,466 tokens of the whole corpus, a total number of 362,580 tokens were not found in any word lists. Having analyzed this list of words, the researchers found that some of these unlisted words enjoy great frequency and range factors; consequently, they are of great importance in devising the final academic word list for the students of health information management. Following Liu and Han's (2015, p. 4) line of research, the existence of these highly

frequent but unlisted words can be partly due to the fact that "First, the AWL does not include some academic words that are commonly used in HIT academic texts and some AWL word families seldom appear in the HIM corpus".

AWL word families accounted for 13.50 percent of all running words in HITC. While some of them have got a quite wide range and frequency of occurrence (DATA, f=13,428), others happened only once or none in entire corpus (IDEALOGY f=1). On the other hand, words which are not in any lists accounted for 16.40percent of all running words (f= 362580). Therefore, establishing an academic word list specific to every field of study is what many scholars are agreed on (Hyland& Tse, 2007; Liu & Han, 2015). To do so, HIMC was analyzed and a total number of 404 word types were found academic based on mentioned criteria (i.e. frequency, range, and specialized occurrence) and they were not listed in GSL or AWL. Having consulted the list with the two field experts, the researchers found 7 of these words (MELLITUS, INSULIN, CARCINOMA, NEOPLASM, MYOCARDIAL, DIABETES, and HEPATITIS) as technical and eliminated them. Accordingly, we came up with the list of 397 words which are not listed in AWL and GSL. The most 30 frequent academic words in HIMC which are out of GSL and AWL are listed below.

Table 8: The First 30 Most Frequent Types (Out of GSL and AWL) in the Present Corpus.

rank	types	range	frequency	%	rank	types	range	frequency	%
1	HEALTHCARE	10	5142	0.23	16	INPATIENT	10	1049	0.04
2	CLINICAL	10	4640	0.20	17	COPYRIGHT	8	935	0.04
3	ELECTRONIC	10	3386	0.15	18	OUTPATIENT	10	730	0.03
4	PHYSICIAN	10	2524	0.11	19	EMERGENCY	9	729	0.03
5	PRIVACY	10	1647	0.07	20	DRUG	10	712	0.03
6	INFORMATICS	10	1641	0.07	21	REIMBURSEMENT	10	642	0.02
7	ONLINE	10	1501	0.06	22	LABORATORY	10	640	0.02
8	SOFTWARE	10	1473	0.06	23	AMBULATORY	10	638	0.02
9	INTERNET	10	1358	0.06	24	SETTINGS	10	623	0.02
10	DATABASE	10	1270	0.05	25	WEB	10	604	0.02
11	MEDICATION	10	1230	0.05	26	CLINICIANS	9	601	0.02
12	MEDICARE	10	1189	0.05	27	CLINIC	10	600	0.02
13	CANCER	10	1183	0.05	28	INTERFACE	10	586	0.02
14	DIAGNOSIS	10	1179	0.05	29	SCANNED	9	573	0.02
15	DISCHARGE	10	1150	0.05	30	STORAGE	10	528	0.02

Finally, the study found out whether the new word list had a better coverage of *health information management* research articles corpus (HIMC), compared to AWL. To answer the last research question (devise a discipline specific academic word list for the students of HIM and verify its coverage on HIMC), the researchers first eliminated GSL word families from the list of most frequent words of HIMC. Then, we combined the most frequent academic words which were coincided in AWL with the most frequent academic words which were not in any lists. The result was Health Information Management Academic Word List (HIMAWL). The first phase leaves us with 451 word families which are listed in AWL and meet the criteria of the study. Then the number of 397 word families (which were not mentioned in any lists) was added. The result was the final list of academic words (encompassing 848 word families) specifically tailored for the students of health information management. In order to test the coverage of the newly devised Health Information Management Academic Word List (HIMAWL) range software was implemented. The results are illustrated in the following table.

Table 9: Coverage of HIMAWL Lexical Items on the HIMC

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
HIMAWL	353407/15.09	848/ 1.77	848
not in the lists	1857059/84.91	55569/98.23	?????
Total	2210466	56418	848

As it is shown in Table 9, the new academic word list (HIMAWL) could improve the coverage rate over Health Information Management Corpus (HIMC) up to 15.09 percent. It is worth mentioning that word families used to devise this HIMAWL were chosen among academic words and every word which was identified as technical one was eliminated from the list. In order to check the coverage of HIMAWL, the following excerpt of HIMC was chosen randomly and analyzed. Bolded words happen in the new word list and can give us a picture of the usage of the newly developed word list over health information management texts.

The health information **directive** has face validity as it integrates the important elements of health information that have been discussed in the literature. From an ethical **perspective**, the **directive** increases **patient** autonomy, facilitates **patient** control over information, fosters openness and transparency and respects several of the ethical principles articulated by Kluge. Whether an information directive would increase or decrease **authoriz**ation for the use of health care information remains unknown and the **topic** for a future **empirical** study. It may exert a differential effect by increasing the use of some forms of information while reducing the **access** for other uses. The **legal status** of such documents is presently unclear, but it is hoped that bringing the **concept** forward for discussion may stimulate **legal** scholarship on this **topic**. How should the directives best be distributed and **administered**? As the health care field becomes increasingly based on information **technology**, it should not be difficult for **individual**s to be able **access** the **directives** either on the **Internet** or on intranets. These issues, as well as the acceptability of the **directive** to **patients**, and the educational **component** that will need to **accompany** it, will be further refined and evaluated **empirically**. The **empirical** evaluation and refinement will **consist** of the following steps. Following the **process** outlined by Berry and Singer for **Cancer Specific** Advance **directives**, key informant **interviews** will be conducted with **stakeholders** involved in ethics, law and **electronic privacy** issues such as **Privacy** Commissioners. This **process** will **create** a **directive** with both face and content validity. **Focus** groups with lay volunteers will provide **input** from the consumer **perspective**. Educational materials will be developed and refined. The **directive** will then be evaluated in a randomized study to determine whether the **directive** can increase **individual**'s sense of empowerment and **security** over their health information.

Out of 299 running words of the above excerpt, more than 46 tokens were coincided with HIMAWL which give us a coverage near to 15% of the whole text. It shows a relative consistency with the results of the present study; furthermore, it does indicate that the present discipline specific academic word list is worth paying due attention by students, teachers, material developers, and practitioners working in the field of health information management.

5. Discussion and Conclusion

Vocabulary has been the main concern of teachers and learners and this issue is even more critical when it is about specific disciplines and fields of study (Nation, 2006; Mudraya, 2006; Ward, 2009). Following West (1953) and Coxhead's (2000) seminal works (the former devised General Service List and the latter came up with Academic World List) many other researchers tried to enrich this venue of research and came up with different discipline specific wordlists (Lei & Liu 2016; Liu & Han, 2015; Todd, 2017; Vongpumivitch, Huang & Chang ,2009).

This study was devoted to compiling a comprehensive list of academic words for the students of health information management. To do so, a corpus (consisting of research articles and textbooks) of 2,210,466 running words was prepared and analyzed using Range software package. The results indicated that AWL and GSL accounted for 88.86 percent of running words in HIMC (the latter word list covers for 70.36 percent of running words, and the former has got a 13.50 percent coverage over HIMC). Having set the aforementioned word selection criteria, the researchers found that the total number of 1006 word families is the most frequent words of the corpus which could be found in GSL and AWL. In the process of research, we found 397 words which were not included in any lists but quite passim through the corpus and met our selection criteria. We added 451 word families from AWL word list (they also met the study's criteria) to the above mentioned words and the result was a new discipline specific word list, specially tailored for

the students of health information management. This word list was used as a base wordlist in Range software and the corpus was analyzed again to show the coverage of the new word list. The results indicated that the new word list can cover for 15.09% of all tokens in the corpus which is an improvement in lexical coverage of the corpus.

It is hoped that the results of the present study can serve as a guide to the students and teachers of Health Information Management and pave the way to a better understanding of the related texts. Also, they could be illuminating for material developers who are trying to design textbooks for this field as it can give the students a rather good command of academic vocabulary knowledge. At last, it is hoped that the present study could contribute to the similar studies in other disciplines.

References

Barnbrook, G. (1996). *Language and computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

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.

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 Specif. Purposes 26*, 502-514.

Cobb, T., Horst, M., 2002. Is there room for an Academic World List in French? Retrieved 15 03 09, from. http://www.lextutor.ca/cv/awl_F.htm.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213 238.

Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly, 45*, 355–362.

Coxhead, A., & Nation, P. (2001). The specialized 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.

Crompton, P. (2013). Characterizing hedging in undergraduate essays by Middle Eastern Students. *The Asian ESP Journal*, 8(2), 55-78.

Gardner, G., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305–327.

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*.

Heatley, A., Nation, P., & Coxhead, A. (2002). RANGE [computer software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

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

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.

Huntley, H. (2005). *Essential academic vocabulary: Mastering the complete academic word list*. Boston: Houghton Mifflin Harcourt.

Hyland, K. (2002) Specificity revisited: how far should we go now? *English for Specific Purposes 21(4): 385-395.*

Hyland, K., & Tse, P. (2007). Is there an 'Academic Vocabulary'? *TESOL Quarterly*, *41*, 235–253.

Juilland, A., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. The Hague: Mouton

Laufer, B., & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Leki, I., & Carson, J. (1994). Students' perception of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly, 28*, 81–101.

Li, Y. Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System, 38*, 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.

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.

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.

Moini, R., & Islamizadeh, I. (2016). Do We Need Discipline-Specific Academic Word Lists? Linguistics Academic Word List (LAWL). *Journal of Teaching Language Skills, 35*(3), 65-90.

Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235-256.

Nagy, W. & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly, 47*(1), 91–108.

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

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63,* 59–82.

Paquot, M. (2007). Towards a productively-oriented Academic Word List. In: Walinski, J., et al. (Eds.), *PALC Proceedings*. Peter Lang, Frankfurt, pp. 127-140.

Samraj, B. (2002). Introductions in research articles: variations across disciplines. *English for Specific Purposes*, *21*(1), 1-17.

Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes, 10,* 189-206.

Sinclair, J.McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. 2005. "Corpus and Text - Basic Principles" in *Developing Linguistic Corpora: a Guide to Good Practice*, ed. M. Wynne. Oxford: Oxbow Books: 1-16. Available online from http://ota.ox.ac.uk/documents/creating/dlc/ .

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31-39.

Townsend, D., & Kiernan, D. (2015). Selecting Academic Vocabulary Words Worth Learning. *The Reading Teacher*, *69*(1), 113–118.

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.

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

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 n on-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28,* 33-41.

Yang, M. (2015). A nursing academic word list. *English for specific purposes, 37*(1), 27-38.