Original Article

A Study on the Most Frequent Academic Words in High Impact Factor English Nursing Journals: A Corpus-based Study

Yadollah Pournia

Department of English Language, Faculty of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran

Abstract

Background: The ability to comprehend a text depends primarily on the knowledge about its words. This study investigated the most frequent words in high impact factor (IF) English nursing journals. **Materials and Methods:** This corpus-based study was conducted on the articles of 13 English nursing journals with an IF of over 0.7 from November 2014 to September 2016. After the typographical errors were corrected and the tokens (running words) in each journal were equalized, the tokens were analyzed using the Range software. Finally, a word list was extracted from the final 2851 articles and 8196,953 tokens to reach the optimal 98% vocabulary coverage. **Results:** A word list consisting of 1081 word families and 3175 word types with 5.24% coverage was extracted, which fulfilled the 98% vocabulary coverage. In other words, the coverage of the 1081 word-family list (5.24%), the coverage of the 1st 3000 English word families (87.55%), proper names, marginal words, compound words, and abbreviations related to the software (3.29%), and the coverage of the new proper names (1.13%), new compounds (0.02%), new abbreviations (0.72%), and letter–number combinations (0.05%) totaled 98%. **Conclusions:** By learning the 1st 3000 English word families and the 1081 word families introduced in this study, a nursing student can comprehend the texts of articles in high IF nursing journals without any considerable help from other resources.

Keywords: *Journal impact factor, nursing, vocabulary*

Introduction

Learning the vocabulary of a language is an essential part of any educational program to learn that language[1,2] because the ability to comprehend a text is mainly dependent on the knowledge about the vocabulary of the text.[1,3] However, teachers of English for specific purposes (ESP) educational programs are often uncertain about what vocabulary their students need to learn.[4,5] To provide learners of ESP with the required language to study texts, they have to be provided with the vocabulary they really need.[6]

In 2000, the Academic Word List (AWL) consisting of 570 word families with a nearly 10% coverage in academic texts was introduced by Coxhead.[7] The word list was obtained from the General Service List provided by West in 1953.[8] The effectiveness of the AWL has sometimes been questioned[5,9] since the word list had the lowest coverage in the texts of sciences including medical sciences with a 9% coverage,[7] and many of the most frequent words in this list rarely appear in the articles of medical sciences.[10] In addition, the effectiveness of the frequent 2000 word-family General Service List from which the AWL was obtained has also been criticized due to the age of the list and the small size of the relevant corpus.[11] Considering these problems of the two lists, some researchers put emphasis on finding frequent discipline-specific vocabulary consistent with the requirements of the discipline.[5,6,12]

Real, comprehensive, public, private, verifiable, and applicable knowledge about humans is obtained through scientific researches,[13] the results of which are often published in the form of articles in scientific journals. In terms of being dynamic, scientific journals are not comparable with textbooks as the contents of many textbooks become outdated even before their publication and they have little educational value for their readers.[14] Journals with high impact factors (IFs) play an important role in providing this dynamic knowledge, and IF is the main and most

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Pournia Y. A study on the most frequent academic words in high impact factor english nursing journals: A corpus-based study. Iranian J Nursing Midwifery Res 2019;24:11-7.

Received: November, 2017. **Accepted:** September, 2018.

Address for correspondence:
Mr. Yadollah Pournia, Department of English Language, Faculty of Medicine, Kamalvand Campus of Lorestan University of Medical Sciences, Khorramabad-Boroujerd Road, Khorramabad, Lorestan, Iran.
E-mail: ypournia@yahoo.com

Access this article online

Website: www.ijnmrjournal.net

DOI: 10.4103/ijnmr.IJNMR_190_17

Quick Response Code:

common indicator to judge the quality of a journal.[15,16] The yearly IF of a scientific journal is the average of citations to all of the articles published in the journal in the previous 2 years[17] and is obtained by dividing the number of citations to the articles published in the journal by all of the articles published in the journal in the previous 2 years.[18]

The importance of nursing scientific journals cannot be ignored. The spread of novel nursing knowledge, which stems from scientific research and is published in nursing journals, is vital for the development of the nursing profession, and this knowledge can have a much higher quality and can be provided much faster compared to what is stated in nursing textbooks.[19] In recent years, due to the emphasis placed by modern medical education on "evidence-based medicine," reading scientific and research papers has been highlighted.[20] In fact, the aim of teaching English to the students of medical sciences in higher levels is mainly to help them in reading and then writing research papers.[21]

Several studies have been conducted on the most frequent vocabulary in nursing texts. Takakubo (2003) analyzed a list of 2650 words at the end of 10 nursing textbooks.[22] Budgell *et al*. introduced a list of 1000 frequent words by studying nursing articles consisting of 250,000 tokens (running words).[23] Mukundan and Jin provided a list of 1004 specialized words by analyzing 3490,417 tokens.[24] Nor Mohamad and Jin extracted a 2000-word list of frequent words by working on 3640,760 tokens in 7 nursing textbooks.[4] In addition, Yang, by doing research on 252 nursing articles consisting of 1006,934 tokens, introduced a list of 676 frequent word families.[25] Most of these studies have investigated nursing vocabulary in nursing textbooks, and none of them has investigated frequent vocabulary in a large number of nursing articles in high IF nursing journals.

Given the importance of vocabulary in learning any language,[1-4] the importance of scientific research,[13] the importance of high IF scientific journals,[15,16] the importance of nursing scientific journals,[19,20] and the need of medical sciences students to English for reading and writing scientific articles in medical journals,[21] this study investigated the most frequent words in high IF (>0.7) English nursing journals and introduced a word list as the High-impact Nursing Academic Word List (HI-NAWL). In addition, the coverage of the 1st common 3000 word families of English and the Coxhead's AWL[7] in nursing scientific journals with an IF over 0.7 was also investigated.

Materials and Methods

This quantitative corpus-based study was conducted from November 2014 to March 2015 (downloading the articles) and from July 2015 to September 2016 on articles downloaded from the ProQuest Nursing and Allied Health Source section of the ProQuest database.[26] The ProQuest database contains thousands of articles published in hundreds of scientific journals and is composed of several smaller databases. The ProQuest Nursing and Allied Health Source includes hundreds of health and nursing scientific journals. First, the general word of "nursing" was searched by ticking the "Full-text" and "Peer-reviewed" tabs so that the found articles would possibly be the most relevant full-text and peer-reviewed articles in nursing. Then, in the journal section which showed the names of the journals in which the found articles had been appeared, journals with an IF higher than 0.7 were selected based on the list of IF in 2012.[27] This cutoff point was chosen based on the number of the journals with an IF in the search results. Most of the journals did not have an IF, and some of the journals with an IF did not have enough full-text articles to be downloaded. IF in nursing journals is generally much lower than that in journals of medicine so that the highest IF in nursing journals (IF = 2.50) belonged to the *Journal of Oncology Nursing Forum*.[28] Therefore, an IF over 0.7 is considered high in nursing. In total, there were 13 nursing journals with an IF over 0.7 in the list of the journals [Table 1]. Subsequently, the articles of each

Table 1: Information related to the journals and the selected articles

Journal	Country	Impact factor	Final number of articles	Final number of words
AAOHN journal	USA	0.85	249	630,668
Clinical Journal of Oncology Nursing	USA	0.91	265	630,551
Critical Care Nurse	USA	0.89	266	630,980
Journal of Gerontological Nursing	USA	0.80	255	630,351
Journal of Neuroscience Nursing	USA	0.75	193	630,521
Journal of Nursing Education	USA	1.13	209	630,627
Journal of Nursing Scholarship	England	1.61	169	630,627
Journal of Psychosocial Nursing and Mental Health Services	USA	0.82	267	630,647
Journal of the American Academy of Nurse Practitioners	USA	0.70	180	630,337
Nursing Ethics	USA	1.21	152	630,238
Oncology Nursing Forum	USA	2.39	195	630,341
Rehabilitation Nursing	England	0.77	188	630,903
The Journal of Continuing Education in Nursing	USA	0.71	263	630,162

AAOHN: American Association of Occupational Health Nursing Journal

journal were arranged and downloaded "most recent first." In fact, the nonprobability consecutive sampling method was used. Downloading was continued to guarantee the collection of the intended sample size of at least 630,000 tokens for each journal and the corpus (collection of texts) size of at least 8 million and 190,000 tokens for the 13 journals. The size of at least 8 million tokens was specified arbitrarily because the present study aimed to be the largest ever done study on nursing vocabulary. The number of the articles needed to cover the words in each journal was not exactly clear. Therefore, the number of the articles downloaded (6000 articles) was more than necessary. The time period of the downloaded articles was between the years 2002 and 2014.

First, the articles were downloaded with Word or PDF formats. The file of each article was converted to Text format and the sections of authors and affiliations, abstracts, acknowledgments, references, tables, and figures were removed from each file. In accordance with the requirements of the software Range[29] used in the study, the necessary corrections were made in each file. Then, the articles of each journal were collected and saved in one Text-format file, and new articles were added to the file to collect the intended tokens and corpus. To conduct corpus-based studies, millions of tokens are required to ensure the availability of large volumes of texts and language samples.[7] The tokens in each journal were counted using the software repeatedly after adding each article to the file of the journal. In this step, the number of the tokens for each journal was more than the minimum required because the number of the tokens would change in the next step due to the required corrections.

In the next step, typographical errors in each journal file were corrected using the Word software. Then the tokens in each journal were recounted to guarantee the presence of 630,000 tokens with a difference of at most 1000 additional tokens in each journal. The additional tokens more than the range of 630,000 + 1000 were removed from the end of the file of each journal. All the files prepared for the 13 nursing journals with roughly equal number of tokens (630,000 + 1000) were analyzed using the software Range, a free software developed by Heatley, Nation, and Coxhead[29] through the 29 word lists of the software including the 25,000 word-family lists of The British National Corpus and the Corpus of Contemporary American English (BNC/COCA)[30] to determine the frequency and range of all the tokens or word families. The 29 lists of the software include the 1st 25,000 word families of English compiled in 25 lists, each with 1000 word families, along with four lists of proper names, marginal words, compound words, and abbreviations. These 29 lists have been prepared based on two big corpora or collections of texts in American and British English. When a text is run on the software, the software measures the frequency and range of each word and the word family to which that word belongs separately based on the words in the 29 lists. When a word does not exist in the 29 lists, the software labels that word as "Not in the Lists." Frequency refers to the number of occurrences of a token or word family, and range refers to the number of the files (journals) in which the token or word family is repeated.[7] In this step, some of the tokens were not in any of the lists of the software and were classified in five separate files of new word families, proper names, compound words, abbreviations, and letter–number combinations in accordance with the requirements of the software.

All the corrected files for the 13 journals consisting of 8196,953 tokens were analyzed through the software through the 34 word lists consisting of the 29 word lists of the software and the 5 word lists prepared in this study to determine the frequency and range of all the tokens or word families.

In the last step, the most frequent word families in the 25,000 word-family lists of BNC/COCA[30] of the software excluding the 1st 3000 word families – which are expected to be learned before entering the university[31] or at most after passing the general English credits before the specialized credits in the university – and the word list of new words prepared in this study were selected using the software.

In this study, the concept of "word family" was used to select the words. In other words, a basic word and its inflected forms and derivations (if any) based on Level 6 of the scale of Bauer and Nation[32] were considered as a word family. For example, the basic word of activate and its following inflected forms and derivations including activated, activates, activating, activation, activator, activators, inactivation, reactivate, reactivated, reactivates, reactivating, reactivation, reactivations, and unactivated are composed of one word family with 15 word types.

If all the above 15 word types appear, for example, 100 times in all the texts, there will be 100 tokens or running words. Therefore, in this example, there will be one word family for the basic word of activate, with 15 word types, and 100 tokens or running words.

To easily select the most frequent words, the 1st three 1000 word-family lists were converted into one 3000 word-family list and the next twenty-two 1000 word-family lists into one 22,000 word-family list. The 25 lists were converted into two lists because of two reasons. First, the conversion would give a better analysis of the 1st 3000 word families of English combined. Second, having one list instead of 22 lists would save a lot of time in selecting the frequent words since the words could be selected once in one list, not 22 times in 22 lists separately. Moreover, the other four lists of the software including proper names, marginal words, compound words, and abbreviations were converted into one list. In

total, there were eight lists including three combined lists of the software and five lists prepared in this study. The most frequent words were selected from the 2nd 22,000 word-family list of the software and the fourth list of new word families prepared in this study.

Two criteria of frequency and range were considered to select the frequent words, and the criterion of the range was before frequency. In other words, words with high frequency would be selected only if they appeared in at least more than half (seven) of the journals.[7] The selected words had to have a range of seven or higher, meaning that they had to be used in at least 7 or more of the 13 nursing journals. The selection of the frequent words was continued to reach 98% coverage, which is the required coverage for the optimal comprehension of language texts without any help from any other sources.[33] In other words, the total of the coverage of the 1st 3000 word families of English and their related compounds, the coverage of the words without any meaning loads such as proper names, marginal words, abbreviations, and letter–number combinations, and the coverage of the selected frequent words had to be 98%.

Due to the large number of the whole words, the selection of the minimum or cutoff frequency required to select the frequent words needed a lot of calculations and trial and errors. After many calculations and trial and errors, it was found that the minimum frequency was 92. In other words, the members of the selected word families had to be repeated at least 92 times in the 13 nursing journals. After removing the word families with a range of <7 using the software of Excel, a list of 1081 word-families was selected as the HI-NAWL, which fulfilled the required 98% coverage. In addition to selecting this list of frequent words, the frequency of the 1st 3000 word families of English and the frequency of the Coxhead's AWL[7] were investigated in the corpus of the 13 nursing journals.

Ethical considerations

This article was derived from a research project (no. 1948-2015) approved by the Research Committee of Lorestan University of Medical Sciences. The study was first approved by the Research Committee of the Faculty of Nursing and Midwifery of the university.

Results

The number of all the articles of the 13 journals in the final stage, after the corrections and classifications, decreased to 2851 articles [Table 1] with 8196,953 tokens. The information of all the tokens of the articles in the 34 word lists is shown in Table 2. According to this table, the 8196,953 tokens of the nursing journals consisted of 82,145 word types and 62,148 word families.

The analysis of all the tokens after converting the 34 word lists into eight word lists is presented in Table 3. According to this table, the first 3000 word families of English (List 1) covered 87.55%, the next 22,000 word families (List 2) 6.52%, other lists of the software including proper names, marginal words, compound words, and abbreviations (List 3) 3.29%, new word families prepared in this study (List 4) 0.71%, new proper names (List 5) 1.13%, new compound words (List 6) 0.02%, new abbreviations (List 7) 0.72%, and letter–number combinations (List 8) 0.05% of all the tokens in the journals.

The selection of the frequent words was performed based on the data presented in Table 3 to reach 98% coverage. The frequent words were selected from the 2nd 22,000 word families of English (List 2) and the new word families prepared in this study (List 4). The results showed that a list of 1081 word families [Supplementary File] consisting of 3175 word types with a range of seven or higher with 5.24% coverage, named the HI-NAWL, fulfilled the required 98% coverage. In other words, the total of the coverage of the 1081 word-family HI-NAWL (5.24%), the coverage of the 1st 3000 word families of English (87.55%), the coverage of other words of the software such as proper names, marginal words, compound words, and abbreviations (3.29%), the coverage of new proper names (1.13%), the coverage of new compound words (0.02%), the coverage of new abbreviations (0.72%), and the coverage of letter–number combinations (0.05) were equal to 98%, which is the required coverage for the optimal comprehension of language texts without any help from any other sources.[33]

The result of the coverage of the Coxhead's AWL[7] showed that 569 word families of this 570 word-family list covered 11.75% of all the tokens in the 13 nursing journals.

Discussion

This study was conducted on 2851 full-text and peer-reviewed articles consisting of 8196,953 tokens in 13 English nursing journals with an IF over 0.7. The number of the words in this study (8196,953 tokens) is much more than the numbers in the studies conducted on nursing words by Takakubo,[22] Budgell *et al*.,[23] Mukundan and Jin,[24] Nor Mohamad and Jin,[4] and Yang.[25] The number is even more than twice the number of the words in the studies by Mukundan and Jin with 3490,417 words[24] and Nor Mohamad and Jin with 3640,760 words,[4] who have used the highest number of words so far. The reason for using a very high number of words in the present study is that, in corpus-based studies, millions of words are required to ensure the availability of large volumes of texts and language samples.[7] Moreover, the number of the articles in this study (2851 articles) is much more than the number of the articles in Budgell *et al*.'s study conducted on the articles in one volume of six nursing journals,[23] and in Yang's study with 252 nursing articles.[25] In addition, instead of using a limited number of nursing textbooks similar to the studies by Takakubo,[22] Mukundan and Jin,[24] and Nor Mohamad and Jin,[4] the present study was conducted on many articles (2851 ones) from 13 HI journals. The reason

Table 2: Place of all the tokens of the 13 nursing journals in the 34 word lists

Word lists*	Number of tokens (%)	Number of word types (%)	Number of word families
1. 1st 1000 word families	5142,072 (62.73)	4744 (5.78)	999
2. 2nd 1000 word families	1113,691 (13.59)	4273 (5.20)	996
3. 3rd 1000 word families	920,779 (11.23)	4285 (5.22)	995
4. 4th 1000 word families	214,812 (2.62)	2906 (3.54)	978
5. 5th 1000 word families	83,837 (1.02)	2227 (2.71)	944
6. 6th 1000 word families	55,073 (0.67)	1918 (2.33)	887
7. 7th 1000 word families	42,078 (0.51)	1532 (1.86)	818
8. 8th 1000 word families	21,059 (0.26)	1298 (1.58)	739
9. 9th 1000 word families	20,163 (0.25)	1063 (1.29)	668
10. 10th 1000 word families	15,628 (0.19)	840 (1.02)	570
11. 11th 1000 word families	18,578 (0.23)	751 (0.91)	512
12. 12th 1000 word families	7585 (0.09)	565 (0.69)	427
13. 13th 1000 word families	9838 (0.12)	525 (0.64)	407
14. 14th 1000 word families	9784 (0.12)	497 (0.61)	363
15. 15th 1000 word families	6647 (0.08)	432 (0.53)	320
16. 16th 1000 word families	4479 (0.05)	371 (0.45)	302
17. 17th 1000 word families	6137 (0.07)	304 (0.37)	259
18. 18th 1000 word families	4889 (0.06)	335 (0.41)	278
19. 19th 1000 word families	2921 (0.04)	242 (0.29)	211
20. 20th 1000 word families	2881 (0.04)	258 (0.31)	220
21. 21st 1000 word families	3383 (0.04)	227 (0.28)	210
22. 22nd 1000 word families	1670 (0.02)	174 (0.21)	168
23. 23rd 1000 word families	1558 (0.02)	174 (0.21)	170
24. 24th 1000 word families	703 (0.01)	136 (0.17)	134
25. 25th 1000 word families	1105 (0.01)	111 (0.14)	96
26. New word families	58,332 (0.71)	7994 (9.73)	6393
27. New proper names	92,836 (1.13)	31,187 (37.97)	31,183
28. New compound words	1566 (0.02)	245 (0.30)	207
29. New abbreviations	59,045 (0.72)	3063 (3.73)	2743
30. Letter-number combinations	3749 (0.05)	630 (0.77)	630
31. Proper names	131,026 (1.60)	7113 (8.66)	6956
32. Marginal words	41,663 (0.51)	73 (0.09)	33
33. Compound words	44,714 (0.55)	1111 (1.35)	814
34. Abbreviations	52,672 (0.64)	541 (0.66)	518
Out of the lists	0 (0.00)	0 (0.00)	0
Total	8196,953 (100)	82,145 (100)	62,148

*The 1st-25th and the 31st-34th word lists belonged to the software and the 26th-30th word lists were classified in this study

Table 3: Place of all the tokens of the 13 nursing journals in the eight word lists

Word lists*	Number of tokens (%)	Number of word types (%)	Number of word families
1. First 3000 word families	7176,542 (87.55)	13,302 (16.19)	2990
2. Next 22,000 word families	534,808 (6.52)	16,886 (20.56)	9681
3. Other lists of the software	270,075 (3.29)	8838 (10.76)	8221
4. New word families	58,332 (0.71)	7994 (9.73)	6393
5. New proper names	92,836 (1.13)	31,187 (37.97)	31,183
6. New compound words	1566 (0.02)	245 (0.30)	207
7. New abbreviations	59,045 (0.72)	3063 (3.73)	2743
8. Letter-number combinations	3749 (0.05)	630 (0.77)	630
Out of the lists	0 (0.00)	0 (0.00)	0
Total	8196,953 (100)	82,145 (100)	62,148

*The 1st-3rd word lists belonged to the software and the 4th-8th word lists were classified in this study

for using numerous articles instead of a limited number of textbooks is that the high number of articles written by numerous authors will solve the problems of individual styles in writing and the high frequency of specific words

by a specific author on a specific subject or field of study.[7] Moreover, in terms of being dynamic, scientific journals are not comparable with textbooks as the contents of many textbooks become outdated even before their publication, and they have little educational value for their readers.[14] All in all, according to an extensive search on the Internet, it seems that the present study is a comprehensive research conducted on the highest number of words and articles in high IF nursing journals.

The results of the present study showed that the 1st 3000 word families of English covered 87.55% of all the 8196,953 running words in the journals. Furthermore, the next 22,000 word families together had a coverage of only 6.52% of all the running words [Tables 2 and 3]. In other words, of 100 words in the nursing journals, approximately 87 words were among the 1st 3000 word families, and only six words were among the next 22,000 word families. Of the 1st 3000 word families, only 10 word families had not been used in the journals. However, of the next 22,000 word families, only 9681 word families had been used, and 12,319 word families had not been used even once in the journals. This result indicates the high frequency of the 1st 3000 word families compared to other word families.[32]

The results of the present study introduced a list of 1081 word families named the HI-NAWL [Supplementary File] consisting of 3175 word types with 5.24% coverage, which fulfilled the required 98% coverage. Unlike other studies conducted on nursing vocabulary, which have selected their frequent words outside the 1st 2000 word families of English,[24,25] the selected words in the present study were outside the 1st 3000 word families of English. Therefore, the coverage of the selected words in the studies by Mukundan and Jin (9.9%)[24] and Yang (13.64%)[25] is higher than that in the present study (5.24%). It should be noted the next 22,000 word families covered only 6.52% of all the tokens in the journals while the 1081 word families in this study covered 5.24% of all the tokens, indicating the importance of the HI-NAWL in this study. Selected academic words for various disciplines including nursing should be outside the 1st 3000 word families of English because these words are expected to be learned before entering the university,[31] or at most after passing the general English credits before passing the specialized credits in the university.

Although the number of the selected words in this study (1081 word families) is higher than the numbers in other studies,[23-25] none of these studies has considered the 98% coverage of words, which is the required coverage for the optimal comprehension of language texts without any help from any other sources.[33] Considering the 98% coverage has resulted in the introduction of a higher number of words in the present study compared to these studies.[23-25]

The results of the frequency of the Coxhead's AWL[7] showed that 569 word families of this 570 word-family list covered 11.75% of all the tokens in the journals, which is a high coverage. The reason is that the words of this list,[7] like the words of the other studies,[4,22-25] are mostly among the first 3000 word families of English with very high frequency.

Downloading the articles, collection of the texts, classifications of the new word families, and the final analyses performed to select the final HI-NAWL in the present study were very time-consuming. Conducting similar studies requires a lot of time, energy, patience, and interest. Furthermore, the full-text articles of some famous nursing journals were not available, and this was one of the limitations of this study. It is recommended that similar studies be conducted on the words of other medical and nonmedical disciplines to extract the required discipline-specific vocabulary.

Conclusion

It can be concluded that by learning the 1st 3000 word families of English and learning the 1081 word-family HI-NAWL introduced in the present study, nursing students and other nursing groups can comprehend nursing texts in high IF nursing journals without any considerable assistance from other sources. Other words including proper names, marginal words, abbreviations, and letter–number combinations, which do not have any specific meaning loads, do not make any considerable problems. The vocabulary extracted in this research accompanied by appropriate passages and exercises can be compiled in educational books, and these books can be used by nursing students and other nursing groups to expand their academic vocabulary in the field of nursing.

Acknowledgments

The researcher appreciates the sincere help by the Faculty of Nursing and Midwifery, Lorestan University of Medical Sciences, for the initial approval, and the Deputy for Research of this university for the final approval and funding of this study (no. 1948-2015).

Financial support and sponsorship

Lorestan University of Medical Sciences

Conflicts of interest

Nothing to declare.

References

1. Hsu W. Measuring the vocabulary of college general English textbooks and English-medium textbooks of business core courses. Electron J Foreign Lang Teach 2009;6:126-49.
2. Csomay E, Petrović M. "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. System 2012;40:305-15.
3. Konstantakis N. Creating a business word list for teaching business English. Estud Lingüíst Inglesa Apl 2007;7:79-102.
4. Nor Mohamad AF, Jin NY. Corpus-based studies on nursing

textbooks. Adv Lang Lit Stud 2013;4:21-8.
5. Alizadeh I, Farjami H. Recounting and fine-tuning academic word list for four academic fields. Iran EFL J 2011;7:48-73.
6. Frazer S. Beyond the Academic Word List: Providing ESP learners with the words they really need. Proceedings of the BAAL Annual Conference; 2008. p. 41-4.
7. Coxhead A. A new academic word list. TESOL Q 2000;34:213-38.
8. West M. A General Service List of English Words. London: Longman, Green; 1953.
9. Tajino A, Dalsky D, Sasao Y. Academic vocabulary reconsidered: An EAP curriculum-design perspective. J Teach Engl Foreign Lang Lit Islam Azad Univ 2009;1:3-21.
10. Chen Q, Ge G. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). Engl Specific Purposes 2007;26:502-14.
11. Browne C. A new general service list: The better mousetrap we've been looking for? Vocabulary Learn Instr 2014;3:1-10.
12. Hyland K, Tse P. Is There an "Academic Vocabulary"? TESOL Q 2007;41;235-53.
13. Masic I. How to search, write, prepare and publish the scientific papers in the biomedical journals. Acta Inform Med 2011;19:68-79.
14. Weatherall DJ, Ledingham JG, Warrell DA. On dinosaurs and medical textbooks. Lancet 1995;346:4-5.
15. Swedlove F. Implications of the impact factor. Can J Occup Ther 2006;73:3-4.
16. Lokker C, Haynes RB, Chu R, McKibbon KA, Wilczynski NL, Walter SD, *et al.* How well are journal and clinical article characteristics associated with the journal impact factor? A retrospective cohort study. J Med Libr Assoc 2012;100:28-33.
17. Moustafa K. The disaster of the impact factor. Sci Eng Ethics 2015;21:139-42.
18. Zucker KJ, Cantor JM. The impact factor: The archives breaks from the pack. Arch Sex Behav 2006;35:7-9.
19. Campbell-Crofts S. The future of nursing journals. Renal Soc Australas J 2012;8:6.
20. Mungra P, Canziani T. Lexicographic studies in medicine: Academic Word List for clinical case histories. Ibérica 2013;25:39-62.
21. Frazer S. Building corpora and compiling pedagogical lists for university medical students. Hiroshima Stud Lang Lang Stud 2013;16:65-88.
22. Takakubo F. Analysis of Vocabulary in English Textbooks for Student Nurses. The Language Teacher; 2003.
23. Budgell B, Miyazaki M, O'Brien M, Perkins R, Tanaka Y. Developing a corpus of the nursing literature: A pilot study. Japan J Nurs Sci 2007;4:21-5.
24. Mukundan J, Jin NY. Development of a technical nursing education word list (NEWL). Int J Innov Engl Lang Teach Res 2012;1:105-24.
25. Yang MN. A nursing academic word list. Engl Specific Purposes 2015;37:27-38.
26. Available from: https://www.search.proquest.com/nahs/results/CDB3C8223266495APQ/1?accountid=35043. [Last accessed on 2015 Mar 31].
27. Available from: http://www.citefactor.org/journal-impact-factor-list-2012.html. [Last accessed on 2014 Nov 01].
28. Oermann MH, Shaw-Kokot J. Impact factors of nursing journals: What nurses need to know. J Contin Educ Nurs 2013;44:293-9.
29. Heatley A, Nation IS, Coxhead A. Range and Frequency Programs; 2002. Available from: http://www.victoria.ac.nz/lals/about/staff/paul-nation. [Last accessed on 2014 Sep 10].
30. Nation IS, Webb S. Researching and Analyzing Vocabulary. Boston: Heinle Cengage Learning; 2011.
31. Hsu W. Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. Lang Teach Res 2013;17:454-84.
32. Bauer L, Nation IS. Word families. Int J Lexicography 1993;6:253-79.
33. Nation IS. How large a vocabulary is needed for reading and listening? Can Modern Lang Rev 2006;63:59-82.

Supplementary File: The word families of the High-impact Nursing Academic Word List (HI-NAWL) arranged from the most to the least frequent

Word families
ET
Medication
Diagnosis
Physician
Cognitive
Ethical
Simulate
Rehabilitate
Chronic
Acute
Oncology
Questionnaire
Dementia
Distress
Impair
Chemotherapy
Ethics
Diabetes
Fatigue
Efficacy
Qualitative
Demography
Diagnose
Preceptor
Adhere
Adverse
Milligram
Cardiac
Prevalent
MS
Protocol
Administer
Surgical
Catheter
Lung
Subscale
Obese
Certify
Dignity
Syndrome
Scenario
Trauma
Incidence
Psychosocial
Therapeutic
Ongoing
Transplant
Tumor
Fluid
Urine
Domain

Contd...

Supplementary File: Contd...

Word families
Inhibit
Neurology
Duration
Nutrition
Prescribe
Oxygen
Enrol
Ventilate
Pediatric
Pulmonary
Diagnostic
Validate
Competence
Cardiovascular
Mentor
Autonomy
Artery
Respirator
Inclusion
Verbal
Alpha
Delirium
Infuse
Hypertension
Regimen
Insulin
Recipient
Palliative
Empower
Dysfunction
Pre
Interprofessional
Nausea
Ward
Regress
Eligible
Geriatric
Expertise
Outpatient
Onset
Toxic
Consistency
Intake
Optimal
Recur
Subjective
Placebo
Minimize
Genome
Inflame
Renal
Glucose
Utilize

Contd...

Supplementary File: Contd...

Word families
Undergraduate
Abnormal
Beneficial
Receptor
Liver
Induce
Bully
Debrief
Milliliter
Postoperative
Hispanic
Interval
Navigate
Seizure
Activate
Retention
Serum
Consensus
Insomnia
Prolong
Pharmacy
Baccalaureate
Cohort
Spouse
Hormone
Incontinent
Vomit
Multidisciplinary
Tailor
Coefficient
Transfuse
Metabolic
Epilepsy
Acid
Sedate
Vaccine
Prescription
Minimal
Anesthesia
Median
Stressor
Malign
Accredit
Periphery
Survivorship
Asthma
Ambulate
Spine
Confidential
Hospice
Intravenous
Pa
Convenience

Contd...

Supplementary File: Contd...

Word families
Mediate
Sodium
Fracture
Literacy
Prostate
Cerebral
Vitamin
Mammogram
Morbid
Bowel
Comorbid
Feasible
Compliance
Informal
Sensory
Ventricular
Bladder
Semester
Classify
Ischemia
Antipsychotic
Kilogram
Discontinue
Donor
Physiological
Transcript
Disparity
Synthesis
Agitate
Antibiotic
Systemic
Esteem
Ulcer
Abdomen
Innovative
Influenza
Holistic
Overweight
Comorbidity
Frail
Advocacy
Coronary
Vascular
Lesion
Warrant
Appraise
Pharmacology
Vaccinate
Burnout
Novice
Verify
Opioid
Schizophrenic

Contd...

Supplementary File: Contd...

Word families
Disseminate
Genomics
Dyspnea
Empirical
Rotate
Cessation
Dilemma
Myocardial
Posttest
Retrospect
Atrium
Wellness
Gerontology
Gastrointestinal
Turnover
Obstruct
Complement
Saturate
Venous
Discomfort
Integrity
Pretest
Prognosis
Diminish
Longitudinal
Edema
Marital
Valve
Parameter
Aspire
Hemorrhage
Fever
Fidelity
Classification
Mutate
Attain
Cognition
Cue
Overview
Invasive
Sepsis
Neuroscience
Calcium
Differentiate
Stigma
Norm
Applicable
Optimize
Tract
Preliminary
Surgeon
Aggression
Airways

Contd...

Supplementary File: Contd...

Word families
Tolerate
Beta
Hygiene
Indirect
Dissatisfy
Relapse
Deficiency
Surveillance
Imperative
Maximize
Coworker
Kidney
Trajectory
Dehydrate
Compress
Refine
Diarrhea
Musculoskeletal
Resuscitate
Doctoral
Rigor
Midwife
Cholesterol
Modality
Allergy
Ovary
Meta
Empathy
Articulate
Anemia
Transcribe
Arthritis
Inherent
Stimulus
Inventory
Pathophysiology
Algorithm
Relevance
Reimburse
Estrogen
Alleviate
Exclusion
Elicit
Exacerbate
Dynamic
Socialize
Enzyme
Secrete
Linear
Concurrent
Menopause
Predominant
Migraine

Contd...

Supplementary File: Contd...

Word families
Wean
Pulse
Obstacle
Continuum
Cellular
Systolic
Compassion
Logistics
Campus
Polar
Deploy
Blog
Credible
Hemoglobin
Pharmaceutical
Nonpharmacologic
Hemodynamic
Susceptible
Diary
Mandate
Deviate
Pneumonia
Encompass
Proactive
Socioeconomic
Teen
Proficient
Replicate
Absent
Fatal
Metabolism
Plasma
Osteoporosis
Infarct
Grief
Dependence
Align
Deteriorate
Tolerance
Fibrosis
Disclose
Rationale
Platelet
Threshold
Coerce
Constipate
Dwell
Integral
Sensation
Trait
Cord
Exemplar
Neuropathy

Contd...

Supplementary File: Contd...

Word families
Etiology
Intracranial
Cervix
Epidemiology
Heparin
Hydrate
Hinder
Comprehend
Sibling
Affiliate
Paradigm
Psychometric
Precaution
Hospitalizations
Metastasis
Maternal
Poster
Postpartum
Anti
Forum
Aerobics
Biopsy
Science fiction
Inhale
Disclosure
Differential
Limb
Dyad
Deprive
Discrepancy
Autistic
Neurologic
Hypotensive
Didactic
Acuity
Caution
Spontaneous
Latino
Physiologic
Scholarship
Beck
Diuretic
Terminal
Intubate
Fibrillate
Lateral
Gastric
Tobacco
Manuscript
Embed
Neuron
Suppress
Entrepreneur

Contd...

Supplementary File: Contd...

Word families
Gait
Intern
Quantify
Bathe
Generalizability
Hierarchy
Hypoglycemia
Adequacy
Predispose
Distract
Causal
Serotonin
Census
Pertain
Cam
Reflex
Colorectal
Aide
Defect
Retrieve
Deem
Analgesic
Audio
Contaminate
Ventricle
Hepatitis
Transitioning
Lipid
Likewise
Diastolic
Attrition
Tertiary
Spectrum
Resilience
Pathology
Massage
Perfusion
Exert
Bypass
Intermittent
Dynamics
Verbatim
Passive
Gynecologic
Bioethics
Saliva
Warfarin
Congruent
Aggregate
Facet
Marrow
Needle
Comparative

Contd...

Supplementary File: Contd...

Word families
Phenomenology
Kin
Compassionate
Calorie
Analytic
Herb
Benzodiazepine
Delineate
Conscience
Hyperglycemia
Disadvantage
Reconcile
Extubation
Preexist
Cumulative
Diffuse
Pedagogy
Anonymous
Electrolyte
Enact
Cytokine
Vein
Intermediate
Radiograph
Thyroid
Atypical
Overload
Nasal
Span
Residue
Mobilize
Conjunction
Conversely
Strive
Pill
Seminar
Mandatory
Multidimensional
Glycemic
Credentialing
Sorrow
Mid
Neonatal
Hepatic
Steroid
Trustworthy
Accord
Bereave
Registry
Antagonist
Autonomous
Convenient
Overlap

Contd...

Supplementary File: Contd...

Word families
Postintervention
Surrogate
Antibody
Dizzy
Sociodemography
Posture
Disposition
Lynch
Thrombosis
Sclerosis
Bachelor
Clot
Cystic
Dental
Congest
Terminology
Counterpart
Cortisol
Diploma
Anticoagulant
Pelvis
Latent
Radiology
Saline
Oxygenate
Vigil
Pertinent
Statin
Esophagus
Pharmacologic
Compatible
Spousal
Tanner
Leisure
Implant
Tablet
Pesticide
Congenital
Licensure
Node
Misconception
Premature
Dopamine
Membrane
Chromosome
Flush
Comply
Adjuvant
Mitigate
Requisite
Appetite
Escalate
Mucus

Contd...

Supplementary File: Contd...

Word families
Futile
Script
Attentive
Infer
Personalize
Dioxide
Marijuana
Lymphoma
Axis
Physiotherapist
Irritate
Robust
Satisfactory
Preceptorial
Tuberculosis
Intrinsic
Dissertation
Confound
Tamoxifen
Psychotropic
Tachycardia
Widow
Cardiopulmonary
Discard
Matrix
Aortic
Scar
Multivariate
Ascertain
Moderated
Lactate
Prophylaxis
Apathy
Magnitude
Kit
Anatomy
Heighten
Reproductive
Exit
Underpin
Pivot
Biomedical
Facial
Necessitate
Anecdote
Psychosis
Subtheme
Multi
Consecutive
Preception
Corticosteroids
Hematologic
Bilateral

Contd...

Supplementary File: Contd...

Word families
Fluctuate
Consortium
Recreation
Entail
Firefight
Yoga
Catheterization
Tomography
Immerse
Graft
Preoperative
Dean
Entity
Sub
ANOVA
Vocation
Composite
Necessity
Inmate
Delude
Click
Psychotic
Genre
Neuromuscular
Chloride
Condom
Subcategory
Anonymity
Definitive
Stratify
Contemplate
Concomitant
Heterogeneous
Repetitive
Physiology
Underestimate
Bundle
Rash
Alternate
Subcutaneous
Stereotype
Waist
Dialysis
Ideation
Irritable
Acculturate
Psychomotor
Postmenopausal
Synergy
Apnea
Resonance
Intact
Assistive

Contd...

Supplementary File: Contd...

Word families
Gauge
Meditate
Categorical
Discourage
Impulse
Detriment
Salient
Liaison
Delete
Dependency
Arena
Endeavour
Osteoarthritis
Withhold
Notify
Compile
Lymph
Nicotine
Perpetrate
Amputate
Contrary
Lens
Portray
Citalopram
Evacuate
Leukemia
Authentic
Viable
Refract
Cannula
Circumference
Beneficent
Coma
Overdose
Latina
Irrigate
Tutor
Normative
Essence
Antecedent
Underscore
Precipitate
Actigraph
Transient
Milieu
Array
Dual
Portable
Redesign
Nonadherence
Neonate
Query
Existential

Contd...

Supplementary File: Contd...

Word families
Semistructure
Ladder
Triage
Mentorship
Circadian
Fetal
Pathological
Bariatric
Sufficiency
Critique
Deplete
Un
Creatinine
Colon
Feedings
Technician
Gamble
Posttrauma
Prone
Packet
Impede
Augment
Agonist
Prognostic
Collegiality
Calendar
Anticoagulation
Triglyceride
Temporal
Certificate
Enrich
Dysphagia
Angiotensin
Adjunct
Biologic
Folic
Orthopedic
Fraction
Nurture
Automate
Sensor
Neuropsychology
Obstetrics
Acetaminophen
Excrete
Ingest
Infrastructure
Rehospitalization
Incise
Intestine
Endocrine
Terminate
Proliferate

Contd...

Supplementary File: Contd...

Word families
Capillary
Plaque
Aspirin
Antigen
Void
Hallucinate
Credentials
Proxy
Reminisce
Stool
Neurosurgery
Numb
Exacerbative
Anterior
Caffeine
Lobe
Oversight
Incur
Metabolite
Remainder
Inverse
Illuminate
Arouse
Sedentary
Tenure
Neuropathic
Involuntary
Hypoxic
Ambivalent
Identical
Rigid
Tolerant
Compel
Endotracheal
Posterior
Histamine
Arrhythmias
Remission
Hurricane
Menstruate
Inclusive
Cardiology
Vagina
Applicant
Behalf
Usage
Inconsistency
Converge
Debilitate
Coagulate
Gland
Assay
Malnutrition

Contd...

Supplementary File: Contd...

Word families
Inter
Reciprocal
Preclude
Enlarge
Lumbar
Bivariate
Mellitus
Contraction
Den
Nonetheless
Fiscal
Interface
Ensue
Acidosis
Pancreatic
Nonuse
Parenteral
Interrater
Potassium
Autoimmune
Benign
Sequelae
Fasts
Whereby
Scarce
Steer
Hermeneutic
Endothelial
Somatic
Glove
Malpractice
Specimen
Regain
Duplicate
Binge
Neurotransmitter
Oximetry
Beneficiary
Biomarker
Postoperation
Closure
Originate
Hybrid
Prophylactic
Mini
Vignette
Underserved
Distal
Defibrillate
Psychotherapy
Covariate
Coherence
Remediate

Contd...

Supplementary File: Contd...

Word families
Titrate
Modulate
Odor
Loop
Organism
Outweigh
Sinus
Basal
Metabolize
Asymptomatic
Thoracic
Hematocrit
Lymphocyte
Necrosis
Guardian
Oppress
Outreach
Solicit
Embody
Interchange
Norepinephrine
Optimism
Formative
Outlook
Artificial
Cultivate
Conversion
Prejudice
Pathogen
Reluctance
Doctorate
Venue
Telehealth
Reciprocity
Morphine
Metaphor
Quiz
Dispense
Rehearse
Gestate
Mistrust
Beverage
Posit
Tangible
Receptive
Conclusive
Pedometer
Overarching
Benchmark
Humiliate
Submission
Linguistic
Numeric

Contd...

Supplementary File: Contd...

Word families
Neurosurgical
Bolus
Carcinoma
Nontraditional
Microscope
Attorney
Inductive
Factorial
Semi
Lipoprotein
Journaling
Foundationalism
Intuitive
Bothersome
Electrode
Embolism
Jersey
Wrist
Sterile
Snack
Rapport
Radiotherapy
Antimicrobial
Reuptake
Genital
Bower
Persuasion
Neutrophil
Suctioning
Pituitary
Analog
Boomer
Nightingale
Enteral
Fetus
Rectal
Deceased
Hoc
Kaiser
Opt
Evoke
Disproportion
Iterate
Anticholinergic
Chemotherapeutic
Decisional
Noncompliance
Avenue
Disconnect
Hemodialysis
Atrophy
Nitrogen
Reed

Contd...

Supplementary File: Contd...

Word families
Viral
Multicenter
Ecological
Respiration
Elective
Intimidate
Twitter
Effusion
Cytotoxic
Adipose
Contraindication
Hematology
Exemplify
Rib
Neuropsychiatry
Autonomic
Neural
Consolidate
Mat
Tolerability
Immigrate
Postdischarge
Pierce
Implicit
Modifiability
Vitality
Perinatal
Nocturnal
Contraindicate
Postgraduate
Mimic
Metropolitan
Paramount
Proximity
Suboptimal
Cautious
Transcend
Reevaluate
Distort
Reflux
Rupture
Moisture
Antihypertension
Cuff
Locus
Nonsignificant
Equity
Ultrasound
Fragile
Elucidate
Jeopardy
Bilingual
Displace

Contd...

Supplementary File: Contd...

Word families
Muscular
Ware
Subjected
Skew
Disfigure
Symptomatology
Increment
Pervasive
Morale
Hallmark
Thread

Contd...

Supplementary File: Contd...

Word families
Uphold
Predetermine
Genotype
Impulsion
Centeredness
Proximal
Aneurysm
Balloon
Respite
Toxin
Theorize
Pencil