

Developing a new academic wordlist for medical purposes – a viable tool for educators

Abstract

Objective: This study aims at developing a list of academic and technical words commonly used in medical research articles. It is conducted in line with the specificity of academic literacy and vocabulary practices in every individual discipline.

Methods: The corpus of this study consisted of 18,462,820 words extracted from 1,784 research articles accessed from three prestigious and widely known journals, i.e., The Lancet, The British Medical Journal (BMJ), and The New England Journal of Medicine (NEJM), published between 2015 and 2019. To analyze the data, the RANGE program as a robust tool for developing viable academic word lists was used.

Results: Our quantitative and qualitative data analysis yielded a final academic wordlist which consisted of 1,003 words, covered 1,972,420 words in the corpus, and accounted for 10.68% of the medical research articles.

Discussion: The high coverage of the extracted academic and technical words provides a reliable source for medical students, medical educators, material designers, and those who are deeply involved in medical English education.

Keywords: medical research article, academic word, technical word, medical students, medical teacher

Khalil Tazik¹

Mahmood Maniati¹

Mohammad Reza

Afshar¹

Somayeh Biparva

Haghighi¹

1 Ahvaz Jundishapur University
of Medical Sciences, Ahvaz,
Iran

1. Introduction

Along with studies conducted on the academic vocabularies over different disciplines, various researchers have examined the academic vocabulary items used in medical science research articles. Chen and Ge [4], for example, examined the coverage of Coxhead's [7] AWL (academic word list) on a corpus of medical research articles with 190,425 running words. Their text analysis revealed that AWL accounted for 10.07% of the medical science research papers, and approximately 292 (51.2%) of the AWL word families were present in most of the research articles. They also examined the dispersion of academic vocabulary items across different rhetorical sections of research articles. They reported that academic vocabulary covered approximately 10% of each section of any given research article. Accordingly, they concluded that AWL word families are important for comprehending medical science research articles and have rhetorical functions in medical research articles. However, they noted that the list requires more modification in order to be regarded as a dependable reference for students of medical sciences.

In another study, Wang et al. [32] compiled a corpus of 1,093,011 running words of medical research articles from different online databases. The final list that they provided included 623 word families that accounted for approximately 12.24% of the total tokens in the corpus. This high frequency was claimed to be a firm evidence

for the generality and comprehensiveness of their medical academic word list. The authors believed that the list can be used as a guide for designing any academic books for medical students, and instructors can refer to the given words in their phases of language teaching.

In a recent study, Lei and Liu [19] questioned the common procedure used for developing academic word lists (i.e., excluding general words from academic words) and claimed that some general high-frequency words can be included in the academic words as they appear to have special meanings in academic contexts. Therefore, they combined Coxhead's [7] procedure with Gardner and Davies's [12] method to develop a new medical academic vocabulary. They reported that the 819 words that they compiled showed higher coverage of medical texts with 53% shorter word count than the previously established medical wordlist by Wang, Liang, and Ge [32].

1.1. Why discipline-specific word lists?

According to Hyland and Tse [14], words behave differently across different contexts and disciplines in terms of their frequencies, collocations, and even meanings. In other words, disciplines follow specific literacy because they have different features and divergent epistemologies [8]. These divergent features were previously discussed by North [25] over the reality-oriented writing in hard sciences and rhetorical-oriented features of soft sciences. Therefore, Hyland and Tse [14] concluded that the fruitful

way for preparing students is “to provide them with an understanding of the features of the discourses they will encounter in their particular courses” (p.251). Along with this recommendation, various discipline-specific studies have been conducted over different disciplines (e.g., engineering wordlist by Hsu [13] and Mudraya [21]; medical word list by Wang et al. [32]; and nursing academic wordlist by Yang [35], to name only a few). What all these studies have in common in terms of their methodology was the exclusion of general service lists (West [33] or BNC most frequent words), following Coxhead's lead [7]. The major motivation underpinning this practice was that the learners grasp highly frequent general words earlier than academic and technical words [19], [24]. Scholars, such as Gardner and Davies [12], however, questioned this assumption and reported some academic words of Coxhead [7] as general words, and thus blurring the borders of academic and general service words. Moreover, Gardner and Davies [12] maintained that a general word such as “interest” has a different meaning in academic texts compared to its meaning in general texts; likewise, some general words are more frequent in academic texts than in general English. Hence, a basic revision is required for developing new academic wordlists for each discipline. Following this method and compiling a corpus of academic and non-academic corpora, Gardner and Davies developed a new academic vocabulary list with a coverage of 13.8% for academic texts, much higher than the coverage reported by Coxhead [7].

1.2. Purpose of the study

The present study was devoted to the development of a new academic word list for medical purposes. It was conducted in line with the specificity of academic literacy and vocabulary practices in every individual discipline. Additionally, new developments in the area of corpus linguistics and recent studies on the development of new GSL (General Service List) motivated us to develop a new word list for medical students.

2. Methods

2.1. Corpus compilation

The corpus of this study consisted of 18,462,820 words extracted from 1,784 research articles. The corpus was set large enough to attain desirable findings and, as Krishnamurthy [16] noted, to detect “finer details of language use” (p.175). The articles were accessed from three prestigious and widely known journals, i.e., *The Lancet*, *The British Medical Journal (BMJ)*, and *The New England Journal of Medicine (NEJM)*, published between 2015 and 2019. The topics covered in these journals included all the specialist areas of the medical research. This advantage helped the researchers to directly focus on the mostly viewed texts for each specialist area. The large number of words included in the corpus and the

wide range of research themes included in the journals made our corpus a representative and reliable source for developing a wordlist for medical purposes.

2.2. Corpus features

1. The corpus composed of all original articles without any format limitation unlike Wang et al. [32] who selected only original articles which had Introduction, Method, Results, and Discussion sections and excluded others.
2. Review articles were included in the corpus since they are critical in medical education with high readership worldwide.
3. The authors and their mother language were not taken into account since the journals were the top peer-reviewed journals with professional language proofreading processes.
4. The author's biodata, tables, figures, pictures, graphs, references, appendices, and the points given in non-English languages were excluded from the corpus.

Data analysis

To analyse the data, the RANGE program downloadable from [https://www.textutor.ca/] was used. Nation [24] stated that this software program is a robust tool for developing viable academic word lists. It has been applied in many previous studies [7], [19], [31], [32], [35]. The program consists of three sub-lists: the two-first base lists belong to West's [33] most frequent general words (GSL) and the third base list includes Coxhead's [7] 570 academic word families. After importing any type of corpus, the program analyses the corpus and ranks the words in accordance with their frequency and inclusion in one of these sub-lists. If the word does not belong to any of these base lists, the word is listed as “not in the list”.

According to the mechanism of the RANGE program and the procedures that researchers follow to develop an academic wordlist, selection of any academic word is based on comparing the lists of words against West's [33] GSL base lists. However, to update the West's [33] GSL, Browne et al. [3] and Brezina and Gablasova [2] developed new GSLs derived from a modern corpus. Subsequently, Eldridge [9] believes that the academic lists established according to these newly developed GSLs contain more discipline-specific words expected by Hyland and Tse [14]. The sources of corpus compilation show that New GSL developed by Brezina and Gablasova [2] is more academic-oriented; therefore, in this study, the researchers preferred Brezina and Gablasova [2] as the GSL reference for refining the academic lists.

The criteria for selecting academic words in this study were:

1. Words that were not included in the GSL [33] and the New General Service List [2] and
2. words that occurred at least 527 times in the entire corpus.

The minimum frequency criterion was duplicated from Coxhead (2000) word selection procedure. Coxhead [7] set the frequency of 100 occurrences in 3.5 million running words, which is about 28.57 times in a million words. This frequency threshold was applied by a large number of studies such as Wang et al. [32], Khani and Tazik [15], and Lei and Liu [19].

After quantitative analysis and identifying the words occurring 527 or more times, in the qualitative phase of analysis, semi-words and proper nouns were omitted from the final list. However, frequently used medical abbreviations and hyphenated words were included in the list. The rationale behind this decision was that in medical pedagogy, instructors are seeking the most relevant and useful vocabulary items to be included in their teaching materials, whether they are common single words or abbreviations and hyphenated words.

Another important issue was comparing the final wordlist against the New General Service List [2] not included in the RANGE program. In this phase of analysis, all words that had a frequency of occurrence of 527 or higher in the entire corpus (whether included in Coxhead's wordlist or not included in any list) were searched for any possible occurrence in the General Service List. Words that appeared in the given list were excluded from the academic words. In this way, all words would be filtered, yielding the most reliable list to be used as academic medical words. Of course, the exclusion of general words was done with great caution. For example, although the noun form of *statistic* is a general word, it was found that its adjective and adverb forms are academic words, occurring 12546 times in the total corpus. Therefore, they were added to the finally reported academic wordlist.

To report the remaining academic words, Coxhead [7] organized her report around the word families, assuming that the most frequent form of the word mirrors the importance of the word itself and its derivational forms, and understanding other forms of the word requires little effort. For instance, *inhibit* is the headword for *inhibited*, *inhibiting*, *inhibition*, *inhibitions*, *inhibitor*, *inhibitors*, *inhibitory*, and *inhibits*. Though highlighting the word family has its own advantages, understanding all the derived forms requires extra knowledge of morphology, which is beyond the students' current level [2]. Furthermore, students seek the best way for decreasing reading comprehension problems, and introducing the most frequent word forms is the most viable response to this challenge. Therefore, the most frequent word types rather than word families were the final reported words.

During the analysis, it was found that some headwords in Coxhead's [7] wordlist did not qualify for the medical academic wordlist; however, careful analysis of their members showed that some family members met the criterion. For example, the word *occurred* only 142 times in its simple form, and based on word family analysis it had to be removed from the final list. However, after analysing its family members, it was observed that the past form of this verb as well as its noun and adjective forms occurred 3275, 1248, 1243 times, respectively.

Hence, we concluded that focusing on the headwords or frequency of word family was not a reliable criterion for making decisions on academic word selection. As a result, we lemmatized all the headwords and checked their frequencies one by one. Along with checking the frequencies, the words were also compared against the GSL [2], and word forms regarded as general words were removed from the list. At the end, those words, which met the criterion but were not listed by Brezina and Gablasova [2] were included as academic words for medical research articles.

After identifying academic words which were shared with Coxhead's [7] AWL, we started analysing the words that met the frequency criterion but grouped under "not in the list". In this phase of analysis, the words were included into academic and technical lists. Distinguishing academic words from technical words was done based on group discussions and consulting experts and dependable dictionaries. In cases of any discrepancies, the words were spotted in the context in which they had been used (the article), and their functions were detected. This process continued until final decisions were made.

3. Results and discussion

A total of 232 out of the 570 words families met the criterion to be selected as medical academic words. However, as noted above, some forms of these words were reported to be among the general word list introduced by Brezina and Gablasova [2]. Moreover, one of the pitfalls of selecting words based on word family is that the frequency of the headwords is the criterion but occurrences of their subheads (other parts of speech) are not considered by the RANGE program. Therefore, in order not to miss these word types and to remove the general words from the academic list, all the qualified word families and word types, which occurred more than 527 times, were lemmatized and checked against the New GSL. After lemmatizing and checking this list against the newly developed GSL [2], 941 word types were removed from the list and 174 word types were selected as academic words for medical RAs (research articles). These words occurred 404,066 times, accounting for 2.18% of the total corpus. This process showed the importance of renewing the academic words for medical purposes and updating studies of such purposes. In fact, it showed that Coxhead's [7] academic word list is unequally valuable for all fields of study, and medical students' needs are different. This need for differentiation is evident from the coverage of Coxhead's [7] wordlist in the entire corpus. Also, it can be argued that updating West's [33] GSL and teaching New GSL by Brezina and Gablasova [2] can be a fruitful prerequisite for comprehending academic texts. Without taking these points into account, the students may consider general words as academic or technical words and look up a huge number of general words in medical dictionaries, which is a time-consuming process.

After identifying the academic words, which were shared with Coxhead [7], it was found that although some words had the frequency requirements for being considered as academic words, they were not among Coxhead's [7] academic words. We put these words in a separate table, and ranked them according to their frequency. Then, we determined whether these words are academic, technical, or general. To start with, a group discussion involving the four authors of this study along with three MDs (doctors of medicine), who had taught English for Medical Purposes (EMP) for some years, was held. During this session, the words were spotted in the context in which they appeared and the participants of the group discussion looked the words up in the dictionaries and commented on the class that the words belonged to. In fact, the functions and uses of the words were the basis of the group decisions. In case of any discrepancies, the final decisions were made by the more experienced members. After the discussion session, out of the total 868 words not found in any list, 39 words (e.g., *patient, cancer, clinical, score, drug, infection, surgery, permission*, etc.) were considered general and removed from the list, 820 word types were considered technical words, and the rest (9 words) were considered academic words. Therefore, the refined list included 829 words, which occurred 1568354 times and accounted for 8.49% of the total corpus. These 829 new words were added to the list of word types shared with Coxhead [7] (see table 1).

The final academic wordlist consisted of 1003 words (see attachment 1), covering 1,972,420 words in the corpus (see table 1). Each word in this list occurred about 106 times per 1000 words. This impressive figure shows the significance of the obtained list, especially compared to the wordlist introduced by Wang et al. [32]. The new medical academic wordlist developed in this study covers 10.68% of the medical science corpus. The efficacy of this coverage, when compared to related studies, becomes clearer. As noted by Laufer [17], given this high coverage, if medical students master these lists of words, they will be able to comprehend the medical texts (at least articles in medical journals).

Coverage across the corpus

The coverage of the wordlist developed in this study was compared to the wordlists established by Wang et al. [32] and Lei and Liu [19] (see table 2).

As can be seen from the table, the coverage of academic and technical words in our study is slightly more than that of Wang et al. [32] and considerably less than that of Lei and Liu [19]. To demonstrate the difference between the three medical wordlists, take this excerpt which is a part of an article from BMJ journal published in 2018. It provides a context for the occurrences and coverage of new medical academic word list in the medical texts. The compatibility of the previously developed wordlists and the differences they have with the given wordlist in this study is clearly provided in this example.

Among the findings for **cancers*** at **specific* sites*** (see table 2), 25-hydroxyvitamin D **concentration*** showed a **significant* inverse association*** with the **risk*** of liver **cancer*** (P for trend=0.006). Further adjustment for **dietary factors*** such as **intake** of total energy, fruits and vegetables, meat, fish and shellfish, isoflavone, green tea, and coffee slightly **attenuated** the **association***, but it remained **significant***, with **hazard** ratios from the second to fourth quarters compared with the first quarter of 0.79 (0.48 to 1.28), 0.71 (0.43 to 1.18), and 0.51 (0.29 to 0.90), **respectively** (P for trend=0.02). In a **subset** of the **population*** with available information, further adjustment for **hepatitis B** and **hepatitis C virus infection* status** and alanine aminotransferase **concentration*** in the **multivariable** model did not appreciably **alter** the **association*** of 25-hydroxyvitamin D with liver **cancer***.

In this 130-words excerpt, academic words found in this study were underlined, Lei and Liu's [19] academic words were bold typed, and Wang et al.'s [32] were marked with an asterisk. As it is shown, 13 (10%) words were underlined, 22 (16.92%) words bold typed, and 16 (12.30%) words were asterisked. It can also be seen that only 4 academic words of this excerpt are repeated in both our list and that of Lei and Liu's [19] study. This difference is related to the word selection procedure used by Lei and Liu [19] and their tendency to ignore technical words. Interestingly, all words identified in Wang et al.'s [32] wordlist were shared with Lei and Liu's [19] wordlist, and all these words were reported by Brezina and Gablasova [2] as general words. If students only rely on the two previously developed wordlists, they will miss precious and influential words, which are highly effective in understanding medical RAs. Therefore, medical students refer to the new GSL [2] before writing or studying RAs. In this way, they could bring this basic knowledge to class and by studying the most frequent academic and technical words presented in this study, they will be equipped with the academic and technical knowledge that is essential for understanding the whole RAs. Of course, we do not underestimate the scientific values of other wordlists, and they can be still used alongside the list developed in this study. In cases of any problems students have with the general or academic uses of words, they can refer to other wordlists and spot the challenging words.

According to the above-mentioned discussions, it could be argued that this wordlist could safely serve the medical students better than previously developed lists because of its time-effectiveness for learning, confidential representation of different ranges of topics and subfields of the medical science, and its higher coverage in terms of both academic and technical words. Another advantage of this wordlist is the exclusion of general words and inclusion of technical words in the analysis. The comparison of the list with other lists shows that the coverage of this list is satisfactorily higher.

Research articles are the most accessible and up-to-date sources of knowledge. All medical educators motivate their students to read these sources in order to keep their knowledge in line with the medical findings and trends

Table 1: Refinement of academic word lists

	Before Refinement			After Refinement		
	*No.	Frequency	Coverage	*No.	Frequency	Coverage
Coxhead's (2000) AWL	1115	840770	4.55%	174	404066	2.18%
Not in the list	868	1855303	10.04%	829	1568354	8.49%
Final Medical Academic & Technical Wordlist	1983	2696073	14.59	1003	1972420	10.68%

*Number of word types which occurred 527 or more times

Table 2: Coverage of academic words and GSL across three corpora

	Academic Words		GSL	GSL + Academic Words
Wang et al. (2008)	1751	10.52%	58.09%	68.61
Lei & Liu (2016)	819	19.44%	54.27%	73.71
This Study	1003	10.68%	*76.06%	86.74%

*Combination of West's (1953) GSL and Brezina and Gablasova's (2015) New GSL

of research in interdisciplinary areas. The list of academic words developed in this study will make the way for easier understanding of the texts, especially for non-native students, and it highlights the most appropriate words for teaching in general and specific courses. Accordingly, medical educators can find this list as a guide for their students and, instead of teaching less frequent words, stick to the most frequent and important academic and technical words that are essential for comprehending research articles or any other medical texts.

4. Educational implications and conclusions

The vocabulary items listed in the new medical wordlist (NMWL) evidence the significant role of this list for the medical students and the members of this discourse community. In line with the discipline-specific studies, the NMWL approves that the words occur and behave differently across different fields of study, generating new concepts for disciplinary academic literacy. This specificity of the word list helps academic material developers to develop materials that are both scientifically scrutinized and experimentally attested, which can be of paramount help to students of medicine. This strand of material development is noted by Murray [22] who acknowledges that by understanding the academic demands of students studying in discrete disciplines, instructors "can research appropriate materials and produce relevant lessons and engaging, and which, therefore, promote learning most effectively" (p.3). Therefore, medical teachers can use the NMWL to provide the most relevant and appropriate teaching materials. Moreover, by using NMWL, medical students will be equipped with a wordlist directly derived from the written scientific medical discourse, enabling them to study research articles independently.

The high coverage of the academic words extracted in the present study should lead to the development of reliable sources for medical students, teachers, material designers, and those who are deeply involved in medical

English education. As noted earlier, the list includes academic words, technical words, most frequent acronyms and abbreviations, and the general words that have special meanings in medical texts. It is recommended that students interested in learning the items of NMWL, first refer to the general words, and then try grasping academic and technical words. In this way, they will be equipped with prerequisite knowledge for understanding medical research articles.

According to the results of the present study, medical English teachers should bring in a medical dictionary to the class for teaching medical academic and technical words if the functions and uses of words in medical contexts are the final aims for comprehending medical texts. Of course, general meanings of the words are also needed to be reviewed, but the major focus should be on the medical uses not the general ones. Since the NMWL includes the high-frequent technical terms, the students will be more encouraged to take the wordlist into serious account and have frequent consultations with it. A question might be posed here about the use of such a wordlist when comprehensive medical dictionaries are already available. The answer lies in the frequency of technical terms in research articles, and that not all technical terms have equal learning value. Students need to learn the most frequent and usable words, rather than cramming their mind with unnecessary items that occur in medical texts once in a blue moon. Moreover, aside from its usefulness, grasping a dictionary of thousands of technical terms is an impossible undertaking, which takes a noticeable amount of time and energy.

Another important issue for teachers to take into account is related to locations of the listed words. Teachers should provide opportunities for medical students to spot word collocations, since researchers such as Hyland and Tse [14] hold that words collocate in distinct ways and having information about these distinct ways is valuable and even mandatory for learning the uses of academic words in special contexts.

Funding

This project [U-98249] was conducted under the support of Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, and approved by the Ethics Committee of this university (Ref. ID: IR.AJUMS.REC:1398.84).

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from <https://doi.org/10.3205/zma001530>

1. Attachment_1.pdf (144 KB)
Academic and technical words in alphabetical order

References

1. Bravo MA, Cervetti A. Teaching vocabulary through text and experience in content areas. In: Graves MF, editor. *Essential readings on vocabulary instructions*. New York: International Reading Association; 2009. p.141-152.
2. Brezina V, Gablasova D. Is there a core general vocabulary? *Introducing the New General Service List*. *Appl Ling*. 2015;36(1):1-22. DOI: 10.1093/applin/amt018
3. Browne C, Culligan B, Phillips J. A New General Service List. 2020. Zugänglich unter/available from: <http://www.newgeneralservicelist.org/>
4. Chen Q, Ge GC. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *Engl Spec Purp*. 2007;26(4):502-514. DOI: 10.1016/j.esp.2007.04.003
5. Chung TM, Nation P. Identifying technical vocabulary. *System*. 2004; 32(2):251-263. DOI: 10.1016/j.system.2003.11.008
6. Chung TM, Nation P. Technical vocabulary in specialised texts. *Read Lang*. 2003;15(2):103.
7. Coxhead A. A new academic word list. *TESOL Quar*. 2000;34(2):213-238. DOI: 10.2307/3587951
8. Durrant P. Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Appl Ling*. 2017;38(2):165-193. DOI: 10.1093/applin/amv011
9. Eldridge J. No, There Isn't an 'Academic Vocabulary,' but...": A Reader Responds to K. Hyland and P. Tse's "Is There an 'Academic Vocabulary'?" *TESOL Quart*. 2008;42(1):109-113. DOI: 10.1002/j.1545-7249.2008.tb00210.x
10. Evans S, Morrison B. Meeting the challenges of English-medium higher education: The first year experience in Hong Kong. *Engl Spec Purp*. 2011;30(3):198-208. DOI: 10.1016/j.esp.2011.01.001
11. Farrell P. Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25. 1990.
12. Gardner D, Davies M. A new academic vocabulary list. *Appl Ling*. 2014;35(3):305-327. DOI: 10.1093/applin/amt015
13. Hsu W. Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *Engl Spec Purp*. 2014;33:54-65. DOI: 10.1016/j.esp.2013.07.001
14. Hyland K, Tse P. Is there an "academic vocabulary"? *TESOL Quart*. 2007;41(2):235-253. DOI: 10.1002/j.1545-7249.2007.tb00058.x
15. Khani R, Tazik K. Towards the development of an academic word list for applied linguistics research articles. *RELC J*. 2013;44(2):209-232. DOI: 10.1177/0033688213488432
16. Krishnamurthy R. Size matters: Creating dictionaries from the world's largest corpus. In: *KOTESOL Proceedings 2001. The Learning Environment: The Classroom and Beyond. Proceedings from the 9th Korea TESOL International Conference; 2001 October 13-14; Seoul, Korea*. p.169.
17. Laufer B. What percentage of lexis is necessary for comprehension? In: Lauren C, Nordman M, editors. *From humans to thinking machines*. Clevedon: Multilingual Matters; 1989. p.316-323.
18. Laufer B, Nation P. A vocabulary-size test of controlled productive ability. *Lang Test*. 1999;16(1):33-51. DOI: 10.1177/026553229901600103
19. Lei L, Liu D. A new medical academic word list: A corpus-based study with enhanced methodology. *J Engl Acad Purp*. 2016;22:42-53. DOI: 10.1016/j.jeap.2016.01.008
20. Martínez IA, Beck SC, Panza CB. Academic vocabulary in agriculture research articles: A corpus-based study. *Engl Spec Purp*. 2009;28(3):183-198. DOI: 10.1016/j.esp.2009.04.003
21. Mudraya O. Engineering English: A lexical frequency instructional model. *Engl Spec Purp*. 2006;25(2):235-256. DOI: 10.1016/j.esp.2005.05.002
22. Murray N. An academic literacies argument for decentralizing EAP provision. *ELT J*. 2016;70(4):435-443. DOI: 10.1093/elt/ccw030
23. Nagy W, Townsend D. Words as tools: Learning academic vocabulary as language acquisition. *Read Res Quart*. 2012;47(1):91-108. DOI: 10.1002/RRQ.011
24. Nation P. How Good Is Your Vocabulary Program? *ESL Mag*. 2001;4(3):22-24. DOI: 10.1177/105382590102400106
25. North S. Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Stud High Educ*. 2005;30(5):517-533. DOI: 10.1080/03075070500249153
26. O'Flynn JA. An Economics Academic Word List (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *J Acad Lang Learn*. 2019;13(1):A28-A87.
27. Pearson J. *Terms in context*. Amsterdam: John Benjamins; 1998. DOI: 10.1075/scl.1
28. Ryan G. Technical vocabulary acquisition through texts. A corpus and a case study in theology classroom. Saarbrücken: Lambert Academic Publishing; 2012.
29. Shaw P. Science research students' composing processes. *Engl Spec Purp*. 1991;10(3):189-206. DOI: 10.1016/0889-4906(91)90024-Q
30. Veenstra J, Sato Y. Creating an Institution-specific Science and Engineering Academic Word List for University Students. *J Asia TEFL*. 2018;15(1):148. DOI: 10.18823/asiatefl.2018.15.1.10.148
31. Vongpumivitch V, Huang JY, Chang YC. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *Engl Spec Purp*. 2009;28(1):33-41. DOI: 10.1016/j.esp.2008.08.003

32. Wang J, Liang SL, Ge GC. Establishment of a medical academic word list. *Engl Spec Purp*. 2008;27(4):442-458. DOI: 10.1016/j.esp.2008.05.003
33. West MP. A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology. London: Longmans, Green & Company; 1953.
34. Xue G, Nation IS. A university word list. *Lang Learn Comm*. 1984;3(2):215-229.
35. Yang MN. A nursing academic word list. *Engl Spec Purp*. 2015;37:27-38. DOI: 10.1016/j.esp.2014.05.003

Please cite as

Tazik K, Maniati M, Afshar MR, Haghighi SB. Developing a new academic wordlist for medical purposes – a viable tool for educators. *GMS J Med Educ*. 2022;39(1):Doc9.
DOI: 10.3205/zma001530, URN: urn:nbn:de:0183-zma0015301

This article is freely available from

<https://doi.org/10.3205/zma001530>

Received: 2021-01-10

Revised: 2021-11-08

Accepted: 2021-11-29

Published: 2022-02-15

Copyright

©2022 Tazik et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Corresponding author:

Mahmood Maniati
Ahvaz Jundishapur University of Medical Sciences, PO
Box: 159, IR-15794-61357 Ahvaz, Iran
maniatim@yahoo.com

Entwicklung einer neuen akademischen Wortliste für medizinische Zwecke – ein brauchbares Werkzeug für Pädagogen

Zusammenfassung

Ziel: Diese Studie zielt darauf ab, eine Liste von akademischen und technischen Wörtern zu erstellen, die in medizinischen Forschungsartikeln häufig verwendet werden. Es wird im Einklang mit den Besonderheiten der akademischen Alphabetisierung und der Wortschatzpraxis in jeder einzelnen Disziplin durchgeführt.

Methoden: Der Korpus dieser Studie bestand aus 18.462.820 Wörtern aus 1.784 Forschungsartikeln, auf die in drei renommierten und weithin bekannten Zeitschriften zugegriffen wurde, nämlich The Lancet, The British Medical Journal (BMJ) und The New England Journal of Medicine (NEJM), veröffentlicht zwischen 2015 und 2019. Zur Analyse der Daten wurde das RANGE-Programm als robustes Werkzeug zur Entwicklung tragfähiger akademischer Wortlisten verwendet.

Ergebnisse: Unsere quantitative und qualitative Datenanalyse ergab eine endgültige akademische Wortliste, die aus 1.003 Wörtern bestand, 1.972.420 Wörter im Korpus abdeckte und 10,68% der medizinischen Forschungsartikel ausmachte.

Diskussion: Die hohe Abdeckung der extrahierten akademischen und technischen Wörter bietet eine zuverlässige Quelle für Medizinstudenten, Medizinpädagogen, Materialdesigner und diejenigen, die sich intensiv mit der medizinischen Englischausbildung beschäftigen.

Schlüsselwörter: medizinischer Forschungsartikel, akademisches Wort, technisches Wort, Medizinstudenten, medizinischer Lehrer

Khalil Tazik¹

Mahmood Maniati¹

Mohammad Reza

Afshar¹

Somayeh Biparva

Haghighi¹

1 Ahvaz Jundishapur University
of Medical Sciences, Ahvaz,
Iran

1. Einleitung

Neben Studien zum akademischen Vokabular verschiedener Disziplinen haben verschiedene Forscher die akademischen Vokabularelemente untersucht, die in medizinwissenschaftlichen Forschungsartikeln verwendet werden. Chen und Ge [4] untersuchten beispielsweise die Abdeckung von Coxheads [7] AWL (Academic Word List) in einem Korpus medizinischer Forschungsartikel mit 190.425 fortlaufenden Wörtern. Ihre Textanalyse ergab, dass AWL 10,07% der medizinwissenschaftlichen Forschungsarbeiten ausmachte und ungefähr 292 (51,2%) der AWL-Wortfamilien in den meisten Forschungsartikeln vorkamen. Sie untersuchten auch die Verteilung akademischer Vokabeln auf verschiedene rhetorische Abschnitte von Forschungsartikeln. Sie berichteten, dass das akademische Vokabular etwa 10% jedes Abschnitts eines bestimmten Forschungsartikels abdeckte. Dementsprechend kamen sie zu dem Schluss, dass AWL-Wortfamilien für das Verständnis von medizinwissenschaftlichen Forschungsartikeln wichtig sind und in medizinischen Forschungsartikeln rhetorische Funktionen haben. Sie stellten jedoch fest, dass die Liste noch weiter modi-

fiziert werden muss, um als zuverlässiges Nachschlagewerk für Medizinstudenten zu gelten.

In einer anderen Studie haben Wang et al. [32] hat ein Korpus von 1.093.011 fortlaufenden Wörtern medizinischer Forschungsartikel aus verschiedenen Online-Datenbanken zusammengestellt. Die endgültige Liste, die sie zur Verfügung stellten, umfasste 623 Wortfamilien, die etwa 12,24% der gesamten Token im Korpus ausmachten. Diese hohe Häufigkeit wurde als fester Beweis für die Allgemeingültigkeit und Vollständigkeit ihrer medizinisch-akademischen Wortliste angesehen. Die Autoren glaubten, dass die Liste als Leitfaden für die Gestaltung von akademischen Büchern für Medizinstudenten verwendet werden kann und die Dozenten in ihren Phasen des Sprachunterrichts auf die angegebenen Wörter zurückgreifen können.

In einer kürzlich durchgeführten Studie hinterfragten Lei und Liu [19] das übliche Verfahren zur Entwicklung akademischer Wortlisten (dh das Ausschließen allgemeiner Wörter aus akademischen Wörtern) und behaupteten, dass einige allgemeine häufig vorkommende Wörter in den akademischen Wörtern enthalten sein können, wie sie erscheinen im akademischen Kontext eine besondere Bedeutung haben. Daher kombinierten sie das Verfahren von Coxhead [7] mit der Methode von Gardner und Davies

[12], um ein neues medizinisch-akademisches Vokabular zu entwickeln. Sie berichteten, dass die von ihnen zusammengestellten 819 Wörter eine höhere Abdeckung medizinischer Texte mit einer um 53% kürzeren Wortzahl aufwiesen als die zuvor erstellte medizinische Wortliste von Wang, Liang und Ge [32].

1.1. Warum fachspezifische Wortlisten?

Laut Hyland und Tse [14] verhalten sich Wörter in verschiedenen Kontexten und Disziplinen in Bezug auf Häufigkeit, Kollokation und sogar Bedeutung unterschiedlich. Mit anderen Worten, Disziplinen folgen einer spezifischen Alphabetisierung, weil sie unterschiedliche Merkmale und divergierende Epistemologien aufweisen [8]. Diese abweichenden Merkmale wurden zuvor von North [25] über das realitätsorientierte Schreiben in den harten Wissenschaften und das rhetorisch orientierte Schreiben in den weichen Wissenschaften diskutiert. Hyland und Tse [14] kamen daher zu dem Schluss, dass der fruchtbare Weg zur Vorbereitung der Studierenden darin besteht, „ihnen ein Verständnis der Merkmale der Diskurse zu vermitteln, denen sie in ihren jeweiligen Kursen begegnen werden“ (S. 251). Zusammen mit dieser Empfehlung wurden verschiedene disziplinspezifische Studien in verschiedenen Disziplinen durchgeführt (z. B. Ingenieur-Wortliste von Hsu [13] und Mudraya [21]; medizinische Wortliste von Wang et al. [32]; und pflegewissenschaftliche Wortliste von Yang [35], um nur einige zu nennen). Allen diesen Studien ist methodisch gemein, dass nach Coxheads Vorbild [7] allgemeine Servicelisten (West [33] oder BNC am häufigsten genannt) ausgeschlossen wurden. Die Hauptmotivation, die dieser Praxis zugrunde lag, war, dass die Lernenden sehr häufige allgemeine Wörter früher als akademische und technische Wörter verstehen [19], [24]. Gelehrte wie Gardner und Davies [12] stellten diese Annahme jedoch in Frage und berichteten einige akademische Wörter von Coxhead [7] als allgemeine Wörter und verwischten damit die Grenzen zwischen akademischen und allgemeinen Dienstwörtern. Darüber hinaus behaupteten Gardner und Davies [12], dass ein allgemeines Wort wie „Interesse“ in akademischen Texten eine andere Bedeutung hat als in allgemeinen Texten. Ebenso kommen einige allgemeine Wörter in akademischen Texten häufiger vor als im allgemeinen Englisch. Daher ist eine grundlegende Überarbeitung für die Entwicklung neuer wissenschaftlicher Wortlisten für jede Disziplin erforderlich.

Nach dieser Methode und der Zusammenstellung eines Korpus aus akademischen und nicht-akademischen Korpora entwickelten Gardner und Davies eine neue akademische Vokabelliste mit einer Abdeckung von 13,8% für wissenschaftliche Texte, viel höher als die von Coxhead berichtete Abdeckung [7].

1.2. Zweck der Studie

Die vorliegende Studie widmete sich der Entwicklung einer neuen wissenschaftlichen Wortliste für medizinische

Zwecke. Es wurde in Übereinstimmung mit den Besonderheiten der akademischen Alphabetisierung und der Wortschatzpraxis in jeder einzelnen Disziplin durchgeführt.

Darüber hinaus haben uns neue Entwicklungen im Bereich der Korpuslinguistik und aktuelle Studien zur Entwicklung neuer GSL (General Service List) motiviert, eine neue Wortliste für Medizinstudenten zu entwickeln.

2. Methoden

2.1. Korpuszusammenstellung

Der Korpus dieser Studie bestand aus 18.462.820 Wörtern, die aus 1.784 Forschungsartikeln extrahiert wurden. Der Korpus war groß genug angelegt, um wünschenswerte Ergebnisse zu erzielen und, wie Krishnamurthy [16] bemerkte, „feinere Details des Sprachgebrauchs“ (S. 175) zu erkennen. Die Artikel wurden in drei renommierten und weithin bekannten Zeitschriften abgerufen, nämlich *The Lancet*, *The British Medical Journal (BMJ)* und *The New England Journal of Medicine (NEJM)*, die zwischen 2015 und 2019 veröffentlicht wurden. Die in diesen Zeitschriften behandelten Themen umfassten alle die Fachgebiete der medizinischen Forschung. Dieser Vorteil half den Forschern, sich direkt auf die meistgesehenen Texte für jedes Fachgebiet zu konzentrieren. Eine große Anzahl von Wörtern im Korpus und die breite Palette von Forschungsthemen in den Zeitschriften machten unser Korpus zu einer repräsentativen und zuverlässigen Quelle für die Entwicklung einer Wortliste für medizinische Zwecke.

2.2. Korpusmerkmale

1. Der Korpus besteht aus allen Originalartikeln ohne jegliche Formatbeschränkung, im Gegensatz zu Wang et al. [32], die nur Originalartikel mit den Abschnitten Einleitung, Methode, Ergebnisse und Diskussion ausgewählt und andere ausgeschlossen haben.
2. Übersichtsartikel wurden in den Korpus aufgenommen, da sie in der medizinischen Ausbildung eine wichtige Rolle spielen und weltweit eine große Leserschaft haben.
3. Die Autoren und ihre Muttersprache wurden nicht berücksichtigt, da es sich bei den Zeitschriften um die besten Peer-Review-Zeitschriften mit professionellen Sprachkorrekturverfahren handelte.
4. Die Personendaten der Autoren, Tabellen, Abbildungen, Bilder, Grafiken, Referenzen, Anhänge und die in anderen Sprachen als Englisch verfassten Punkte wurden aus dem Korpus ausgeschlossen.

Datenanalyse

Zur Analyse der Daten wurde das herunterladbare RANGE-Programm [<https://www.lexutor.ca/>] verwendet. Nation [24] gab an, dass dieses Softwareprogramm ein robustes

Werkzeug zur Entwicklung brauchbarer akademischer Wortlisten ist. Es wurde in vielen früheren Studien angewendet [7], [19], [31], [32], [35]. Das Programm besteht aus drei Unterlisten: Die ersten beiden Basislisten gehören zu Wests [33] am häufigsten verwendeten allgemeinen Wörtern (GSL) und die dritte Basisliste enthält die 570 akademischen Wortfamilien von Coxhead [7]. Nach dem Importieren eines beliebigen Korpusstyps analysiert das Programm den Korpus und ordnet die Wörter entsprechend ihrer Häufigkeit und Aufnahme in eine dieser Unterlisten ein. Wenn das Wort zu keiner dieser Basislisten gehört, wird das Wort als „nicht in der Liste“ aufgeführt. Gemäß dem Mechanismus des RANGE-Programms und den Verfahren, die Forscher zur Entwicklung einer akademischen Wortliste befolgen, basiert die Auswahl jedes akademischen Wortes auf dem Vergleich der Wortlisten mit West's [33] GSL-Basislisten. Um jedoch [33] GSL des Westens zu aktualisieren, haben Browne et al. [3] und Brezina und Gablasova [2] entwickelten neue GSLs, die von einem modernen Korpus abgeleitet wurden. Anschließend glaubt Eldridge [9], dass die nach diesen neu entwickelten GSLs erstellten akademischen Listen mehr disziplinspezifische Wörter enthalten, die von Hyland und Tse [14] erwartet werden. Die Quellen der Korpuszusammenstellung zeigen, dass die von Brezina und Gablasova [2] entwickelte New GSL eher akademisch-orientiert ist. Daher bevorzugten die Forscher in dieser Studie Brezina und Gablasova [2] als GSL-Referenz für die Verfeinerung der akademischen Listen.

Die Kriterien für die Auswahl akademischer Wörter in dieser Studie waren:

1. Wörter, die nicht in der GSL [33] und der Neuen Allgemeinen Leistungsliste [2] enthalten waren und
2. Wörter, die im gesamten Korpus mindestens 527 Mal vorgekommen sind.

Das Kriterium der minimalen Häufigkeit wurde aus dem Wortauswahlverfahren von Coxhead (2000) dupliziert. Coxhead [7] hat die Häufigkeit von 100 Vorkommen in 3,5 Millionen laufenden Wörtern festgelegt, was etwa 28,57 Mal in einer Million Wörtern entspricht. Diese Häufigkeitsschwelle wurde von einer Vielzahl von Studien wie Wang et al. [32], Khani und Tazik [15] und Lei und Liu [19].

Nach der quantitativen Analyse und der Identifizierung der Wörter, die 527 oder mehr vorkommen, wurden in der qualitativen Phase der Analyse Halbwörter und Eigennamen aus der endgültigen Liste weggelassen. Häufig verwendete medizinische Abkürzungen und Wörter mit Bindestrich wurden jedoch in die Liste aufgenommen. Der Grund für diese Entscheidung war, dass Lehrende in der Medizinpädagogik nach den relevantesten und nützlichsten Vokabeln suchen, die in ihre Lehrmaterialien aufgenommen werden sollten, seien es gebräuchliche einzelne Wörter oder Abkürzungen und Wörter mit Bindestrich.

Ein weiteres wichtiges Thema war der Vergleich der endgültigen Wortliste mit der New General Service List [2], die nicht im RANGE-Programm enthalten ist. In dieser

Analysephase wurden alle Wörter mit einer Häufigkeit von 527 oder höher im gesamten Korpus (ob in Coxheads Wortliste enthalten oder in keiner Liste enthalten) nach einem möglichen Vorkommen in der General Service List durchsucht. Wörter, die in der angegebenen Liste auftauchen, wurden von den akademischen Wörtern ausgeschlossen. Auf diese Weise würden alle Wörter gefiltert, um die zuverlässigste Liste zu erhalten, die als akademische medizinische Wörter verwendet werden kann. Der Ausschluss allgemeiner Wörter erfolgte natürlich mit großer Vorsicht. Obwohl beispielsweise die Nomenform der Statistik ein allgemeines Wort ist, wurde festgestellt, dass ihre Adjektiv- und Adverbformen akademische Wörter sind, die im Gesamtkorpus 12546 Mal vorkommen. Daher wurden sie der endgültigen gemeldeten akademischen Wortliste hinzugefügt.

Um die verbleibenden akademischen Wörter zu berichten, organisierte Coxhead [7] ihren Bericht um die Wortfamilien herum, wobei angenommen wurde, dass die häufigste Form des Wortes die Bedeutung des Wortes selbst und seiner Ableitungsformen widerspiegelt und das Verständnis anderer Wortformen wenig Aufwand erfordert. Inhibieren ist beispielsweise das Stichwort für inhibiert, inhibierend, inhibierend, inhibierend, inhibierend, inhibierend und inhibiert. Obwohl das Hervorheben der Wortfamilie seine eigenen Vorteile hat, erfordert das Verständnis aller abgeleiteten Formen zusätzliche Kenntnisse der Morphologie, die über das derzeitige Niveau der Schüler hinausgehen [2]. Darüber hinaus suchen die Schüler nach dem besten Weg, um Leseverständnisprobleme zu verringern, und die Einführung der häufigsten Wortformen ist die praktikabelste Antwort auf diese Herausforderung. Daher waren die am häufigsten gemeldeten Wörter die häufigsten Wortarten und nicht die Wortfamilien.

Während der Analyse wurde festgestellt, dass einige Stichwörter in Coxheads [7] Wortliste nicht für die medizinisch-akademische Wortliste geeignet waren. Eine sorgfältige Analyse ihrer Mitglieder zeigte jedoch, dass einige Familienmitglieder das Kriterium erfüllten. Beispielsweise kam das Wort in seiner einfachen Form nur 142 Mal vor und musste aufgrund der Wortfamilienanalyse aus der endgültigen Liste gestrichen werden. Nach der Analyse seiner Familienmitglieder wurde jedoch beobachtet, dass die Vergangenheitsform dieses Verbs sowie seine Nomen- und Adjektivformen 3275, 1248 bzw. 1243 Mal vorkamen. Daher kamen wir zu dem Schluss, dass die Fokussierung auf die Stichwörter oder die Häufigkeit der Wortfamilie kein zuverlässiges Entscheidungskriterium für die wissenschaftliche Wortauswahl ist. Als Ergebnis haben wir alle Stichwörter lemmatisiert und deren Häufigkeiten einzeln überprüft. Neben der Überprüfung der Häufigkeiten wurden die Wörter auch mit der GSL [2] verglichen und als allgemeine Wörter angesehene Wortformen aus der Liste gestrichen. Am Ende wurden diejenigen Wörter, die das Kriterium erfüllten, aber nicht von Brezina und Gablasova [2] aufgeführt wurden, als akademische Wörter für medizinische Forschungsartikel aufgenommen.

Nachdem wir akademische Wörter identifiziert hatten, die mit Coxheads [7] AWL geteilt wurden, begannen wir mit der Analyse der Wörter, die das Häufigkeitskriterium erfüllten, aber unter „nicht in der Liste“ gruppiert waren. In dieser Analysephase wurden die Wörter in akademische und technische Listen aufgenommen. Die Unterscheidung von akademischen Wörtern von technischen Wörtern erfolgte auf der Grundlage von Gruppendiskussionen und der Beratung von Experten und zuverlässigen Wörterbüchern. Bei Unstimmigkeiten wurden die Wörter in dem Kontext, in dem sie verwendet wurden (dem Artikel), erkannt und ihre Funktion erkannt. Dieser Prozess dauerte bis zur endgültigen Entscheidung.

3. Ergebnisse und Diskussion

Insgesamt erfüllten 232 der 570 Wortfamilien das Kriterium, als medizinisch-akademische Wörter ausgewählt zu werden. Jedoch, wie oben erwähnt, wurde berichtet, dass einige Formen dieser Wörter zu der allgemeinen Wortliste gehören, die von Brezina und Gablasova eingeführt wurde [2]. Darüber hinaus besteht eine der Fallstricke bei der Auswahl von Wörtern auf der Grundlage der Wortfamilie darin, dass die Häufigkeit der Stichwörter das Kriterium ist, aber das Vorkommen ihrer Zwischenüberschriften (andere Wortarten) vom RANGE-Programm nicht berücksichtigt wird. Um diese Wortarten nicht zu übersehen und die allgemeinen Wörter aus der akademischen Liste zu entfernen, wurden daher alle qualifizierten Wortfamilien und Wortarten, die mehr als 527 Mal vorkamen, lemmatisiert und gegen die Neue GSL geprüft. Nach Lemmatisierung und Überprüfung dieser Liste mit der neu entwickelten GSL [2] wurden 941-Worttypen aus der Liste entfernt und 174-Worttypen als akademische Wörter für medizinische RAs (Forschungsartikel) ausgewählt. Diese Wörter kamen 404.066 Mal vor, was 2,18% des Gesamtkorpus ausmachte. Dieser Prozess hat gezeigt, wie wichtig es ist, die akademischen Wörter für medizinische Zwecke zu erneuern und Studien zu solchen Zwecken zu aktualisieren. Tatsächlich zeigte sich, dass die akademische Wortliste von Coxhead [7] für alle Studienrichtungen ungleich wertvoll ist und die Bedürfnisse der Medizinstudenten unterschiedlich sind. Dieser Differenzierungsbedarf wird durch die Abdeckung der Wortliste von Coxhead [7] im gesamten Korpus deutlich. Auch kann argumentiert werden, dass die Aktualisierung von Wests [33] GSL und das Lehren der Neuen GSL von Brezina und Gablasova [2] eine fruchtbare Voraussetzung für das Verständnis wissenschaftlicher Texte sein können. Ohne diese Punkte zu berücksichtigen, können die Studierenden allgemeine Wörter als akademische oder technische Wörter betrachten und eine große Anzahl allgemeiner Wörter in medizinischen Wörterbüchern nachschlagen, was ein zeitaufwändiger Prozess ist.

Nach der Identifizierung der akademischen Wörter, die mit Coxhead [7] geteilt wurden, wurde festgestellt, dass einige Wörter zwar die Häufigkeitsanforderungen hatten, um als akademische Wörter angesehen zu werden, sie

jedoch nicht zu den akademischen Wörtern von Coxhead [7] gehörten. Wir haben diese Wörter in eine separate Tabelle aufgenommen und sie nach ihrer Häufigkeit geordnet. Dann haben wir festgestellt, ob diese Wörter akademisch, technisch oder allgemein sind. Zu Beginn fand eine Gruppendiskussion mit den vier Autoren dieser Studie sowie drei MDs (Doctors of Medicine) statt, die einige Jahre Englisch für medizinische Zwecke (EMP) unterrichtet hatten. Während dieser Sitzung wurden die Wörter in dem Kontext entdeckt, in dem sie auftraten, und die Teilnehmer der Gruppendiskussion schlugen die Wörter in den Wörterbüchern nach und kommentierten die Klasse, zu der die Wörter gehörten. Tatsächlich waren die Funktionen und Verwendungen der Wörter die Grundlage der Gruppenentscheidungen. Bei Unstimmigkeiten wurden die endgültigen Entscheidungen von den erfahreneren Mitgliedern getroffen. Nach der Diskussionsitzung wurden von den insgesamt 868 Wörtern, die in keiner Liste gefunden wurden, 39 Wörter (z. B. Patient, Krebs, Klinik, Punktzahl, Medikament, Infektion, Operation, Erlaubnis usw.) 820-Wort-Typen wurden als technische Wörter und der Rest (9 Wörter) als akademische Wörter angesehen. Daher umfasste die verfeinerte Liste 829 Wörter, die 1568354 Mal vorkamen und 8,49% des Gesamtkorpus ausmachten. Diese 829 neuen Wörter wurden der Liste der mit Coxhead geteilten Wortarten [7] hinzugefügt (siehe Tabelle 1).

Die endgültige akademische Wortliste bestand aus 1003 Wörtern (siehe Anhang 1, nur auf Englisch verfügbar) und deckte 1.972.420 Wörter im Korpus ab (siehe Tabelle 1). Jedes Wort in dieser Liste kam etwa 106-mal pro 1000 Wörter vor. Diese beeindruckende Zahl zeigt die Bedeutung der erhaltenen Liste, insbesondere im Vergleich zu der von Wang et al. eingeführten Wortliste [32]. Die in dieser Studie entwickelte neue medizinisch-akademische Wortliste deckt 10,68% des medizinisch-wissenschaftlichen Korpus ab. Die Wirksamkeit dieser Abdeckung wird im Vergleich zu verwandten Studien deutlicher. Wie Laufer [17] feststellte, können Medizinstudenten angesichts dieser hohen Abdeckung, wenn sie diese Wortlisten beherrschen, die medizinischen Texte (zumindest Artikel in medizinischen Fachzeitschriften) verstehen.

Abdeckung im gesamten Korpus

Die Abdeckung, der in dieser Studie entwickelten Wortlisten, wurde mit den Wortlisten von Wang et al. [32] und Lei und Liu [19] verglichen (siehe Tabelle 2).

Wie aus der Tabelle ersichtlich ist, deckt unsere Studie akademische und technische Wörter etwas mehr ab als die von Wang et al. [32] und deutlich weniger als bei Lei und Liu [19]. Um den Unterschied zwischen den drei medizinischen Wortlisten zu demonstrieren, nehmen Sie diesen Auszug, der Teil eines Artikels aus der BMJ-Zeitschrift ist, der 2018 veröffentlicht wurde. Er bietet einen Kontext für das Vorkommen und die Berichterstattung über neue medizinisch-akademische Wortlisten in den medizinischen Texten. Die Kompatibilität der zuvor entwickelten Wortlisten und die Unterschiede, die sie mit der

Tabelle 1: Verfeinerung akademischer Wortlisten

	Vor der Verfeinerung			Vor der Verfeinerung		
	*Nein.	Frequenz	Abdeckung	*Nein.	Frequenz	Abdeckung
Coxhead's (2000) AWL	1115	840770	4.55%	174	404066	2.18%
Nicht in der Liste	868	1855303	10.04%	829	1568354	8.49%
Finale medizinische, akademische und technische Wortliste	1983	2696073	14.59	1003	1972420	10.68%

* Anzahl der Wortarten, die 527 oder öfter vorgekommen sind

Tabelle 2: Abdeckung von akademischen Wörtern und GSL in drei Korpora

	Akademische Wörter		GSL	GSL + Akademische Wörter
Wang et al. (2008)	1751	10.52%	58.09%	68.61
Lei & Liu (2016)	819	19.44%	54.27%	73.71
Diese Studie	1003	10.68%	*76.06%	86.74%

* Kombination aus Wests (1953) GSL und Brezina und Gablasovas (2015) New GSL

gegebenen Wortliste in dieser Studie haben, wird in diesem Beispiel deutlich gemacht.

Unter den Ergebnissen für **Krebs*** an **bestimmten* Stellen*** (siehe Tabelle 2) zeigte die **Konzentration** von 25-Hydroxyvitamin D* eine **signifikante* inverse Assoziation*** mit dem **Risiko*** für **Leberkrebs*** (P für Trend=0,006). Eine weitere Anpassung an **Ernährungsfaktoren*** wie die **Aufnahme** von Gesamtenergie, Obst und Gemüse, Fleisch, Fisch und Schalentiere, Isoflavon, grüner Tee und Kaffee **schwächte** die **Assoziation** leicht ab*, blieb jedoch **signifikant***, mit **Hazard Ratio** von der zweiten bis zu a viertes Quartal im Vergleich zum ersten Quartal von 0,79 (0,48 bis 1,28), 0,71 (0,43 bis 1,18) **bzw.** 0,51 (0,29 bis 0,90) (P für Trend=0,02). In einer **Untergruppe** der **Bevölkerung*** mit verfügbaren Informationen veränderte eine weitere Anpassung des Status einer **Hepatitis-B-** und **Hepatitis-C-Virusinfektion*** und der Alanin-Aminotransferase-**Konzentration*** im **multivariablen** Modell die **Assoziation*** von 25-Hydroxyvitamin D mit **Leberkrebs*** nicht **merklich**.

In diesem 130-Wörter-Auszug wurden die in dieser Studie gefundenen akademischen Wörter unterstrichen, die akademischen Wörter von Lei und Liu [19] fett gedruckt und die von Wang et al. [32] mit einem Sternchen gekennzeichnet. Wie gezeigt, wurden 13 (10%) Wörter unterstrichen, 22 (16,92%) Wörter fett geschrieben und 16 (12,30%) Wörter mit einem Sternchen versehen. Es ist auch zu erkennen, dass sowohl in unserer Liste als auch in der Studie von Lei und Liu [19] nur 4 akademische Wörter dieses Auszuges wiederholt werden. Dieser Unterschied hängt mit dem von Lei und Liu [19] verwendeten Wortauswahlverfahren und ihrer Tendenz zusammen, technische Wörter zu ignorieren. Interessanterweise wurden alle Wörter, die in der Wortliste von Wang et al. [32] identifiziert wurden, mit der Wortliste von Lei und Liu [19] geteilt, und alle diese Wörter wurden von Brezina und Gablasova [2] als allgemeine Wörter gemeldet.

Wenn sich die Schüler nur auf die beiden zuvor entwickelten Wortlisten verlassen, werden ihnen wertvolle und einflussreiche Wörter entgehen, die beim Verständnis

medizinischer RAs sehr effektiv sind. Daher beziehen sich Medizinstudenten auf die neue GSL [2], bevor sie RAs schreiben oder studieren. Auf diese Weise können sie dieses Grundwissen in den Unterricht einbringen und werden durch das Studium der häufigsten akademischen und technischen Wörter, die in dieser Studie vorgestellt werden, mit dem akademischen und technischen Wissen ausgestattet, das für das Verständnis der gesamten RAs unerlässlich ist. Natürlich unterschätzen wir die wissenschaftlichen Werte anderer Wortlisten nicht und können dennoch neben der in dieser Studie entwickelten Liste verwendet werden. Bei Problemen mit der allgemeinen oder akademischen Verwendung von Wörtern können die Schüler auf andere Wortlisten zurückgreifen und die herausfordernden Wörter erkennen.

Nach den oben genannten Diskussionen könnte argumentiert werden, dass diese Wortliste den Medizinstudenten aufgrund ihrer Zeiteffektivität für das Lernen, der vertraulichen Darstellung verschiedener Themenbereiche und Teilgebiete der Medizinwissenschaft sicher besser dienen könnte als zuvor entwickelte Listen seine höhere Abdeckung in Bezug auf akademische und technische Wörter. Ein weiterer Vorteil dieser Wortliste ist der Ausschluss allgemeiner Wörter und die Einbeziehung von Fachwörtern in die Analyse. Der Vergleich der Liste mit anderen Listen zeigt, dass die Abdeckung dieser Liste zufriedenstellend höher ist.

Forschungsartikel sind die zugänglichsten und aktuellsten Wissensquellen. Alle Medizinpädagogen motivieren ihre Studierenden, diese Quellen zu lesen, um ihr Wissen mit den medizinischen Erkenntnissen und Forschungstrends in interdisziplinären Bereichen in Einklang zu bringen. Die in dieser Studie entwickelte Liste wissenschaftlicher Wörter erleichtert das Verständnis der Texte, insbesondere für nicht-muttersprachliche Studierende, und hebt die am besten geeigneten Wörter für den Unterricht in allgemeinen und spezifischen Kursen hervor. Dementsprechend können Medizinpädagogen diese Liste als Orientierungshilfe für ihre Studierenden finden und sich, anstatt weniger häufige Wörter zu lehren, bei den häufigsten

und wichtigsten wissenschaftlichen und technischen Wörtern festhalten, die für das Verständnis von Forschungsartikeln oder anderen medizinischen Texten unerlässlich sind.

4. Pädagogische Implikationen und Schlussfolgerungen

Die in der neuen medizinischen Wortliste (NMWL) aufgeführten Wortschatzelemente belegen die bedeutende Rolle dieser Liste für die Medizinstudenten und die Mitglieder dieser Diskursgemeinschaft. Im Einklang mit den fachspezifischen Studien bestätigt die NMWL, dass die Wörter in verschiedenen Studienrichtungen unterschiedlich vorkommen und sich verhalten, wodurch neue Konzepte für die disziplinäre Wissenschaftskompetenz generiert werden.

Diese Spezifität der Wortliste hilft akademischen Materialentwicklern, sowohl wissenschaftlich geprüfte als auch experimentell belegte Materialien zu entwickeln, die für Medizinstudenten eine große Hilfe sein können. Dieser Strang der Materialentwicklung wird von Murray [22] angemerkt, der anerkennt, dass Dozenten durch das Verständnis der akademischen Anforderungen von Studenten, die in einzelnen Disziplinen studieren, „geeignete Materialien recherchieren und relevante und ansprechende Lektionen erstellen können, die daher das Lernen am effektivsten fördern“. (S. 3). Daher können Medizinlehrer die NMWL nutzen, um die relevantesten und geeignetsten Lehrmaterialien bereitzustellen. Darüber hinaus wird den Medizinstudierenden durch die Nutzung der NMWL eine direkt aus dem schriftlichen wissenschaftlichen medizinischen Diskurs abgeleitete Wortliste zur Verfügung gestellt, die es ihnen ermöglicht, Forschungsartikel selbstständig zu studieren.

Die hohe Abdeckung der in der vorliegenden Studie extrahierten akademischen Wörter sollte zur Entwicklung zuverlässiger Quellen für Medizinstudenten, Lehrer, Materialdesigner und diejenigen führen, die sich intensiv mit der medizinischen Englischausbildung beschäftigen. Wie bereits erwähnt, enthält die Liste akademische Wörter, technische Wörter, die häufigsten Akronyme und Abkürzungen sowie die allgemeinen Wörter, die in medizinischen Texten eine besondere Bedeutung haben. Schülern, die daran interessiert sind, die Elemente der NMWL zu lernen, wird empfohlen, sich zuerst auf die allgemeinen Wörter zu beziehen und dann zu versuchen, akademische und technische Wörter zu erfassen. Auf diese Weise werden sie mit den notwendigen Kenntnissen ausgestattet, um medizinische Forschungsartikel zu verstehen.

Nach den Ergebnissen der vorliegenden Studie sollten Medizinenglischlehrer ein medizinisches Wörterbuch zum Lehren medizinisch-wissenschaftlicher und technischer Wörter in den Unterricht mitbringen, wenn Funktionen und Verwendungen von Wörtern im medizinischen Kontext das Endziel für das Verständnis medizinischer Texte sind. Natürlich müssen auch die allgemeinen Bedeutungen der Wörter überprüft werden, aber der Schwerpunkt

sollte auf den medizinischen Anwendungen liegen, nicht auf den allgemeinen. Da die NMWL die hochfrequenten Fachbegriffe beinhaltet, werden die Studierenden eher ermutigt, die Wortliste ernst zu nehmen und sich häufig damit auseinanderzusetzen. An dieser Stelle könnte die Frage nach der Verwendung einer solchen Wortliste gestellt werden, wenn bereits umfassende medizinische Wörterbücher zur Verfügung stehen. Die Antwort liegt in der Häufigkeit von Fachbegriffen in Forschungsartikeln und dass nicht alle Fachbegriffe den gleichen Lernwert haben. Die Schüler müssen die häufigsten und nützlichsten Wörter lernen, anstatt sich mit unnötigen Gegenständen zu vollstopfen, die in medizinischen Texten einmal in einem blauen Mond vorkommen. Abgesehen von seiner Nutzlosigkeit ist das Erfassen eines Wörterbuchs mit Tausenden von Fachbegriffen zudem ein unmögliches Unterfangen, das einen beträchtlichen Zeit- und Energieaufwand erfordert.

Ein weiteres wichtiges Thema, das Lehrer berücksichtigen sollten, betrifft die Position der aufgelisteten Wörter. Lehrer sollten Medizinstudenten die Möglichkeit bieten, Wortkollokationen zu erkennen, da Forscher wie Hyland und Tse [14] der Ansicht sind, dass Wörter auf unterschiedliche Weise kollokieren und Informationen über diese unterschiedlichen Arten wertvoll und sogar obligatorisch sind, um die Verwendung akademischer Wörter in zu lernen besondere Kontexte.

Förderung

Dieses Projekt [U-98249] wurde mit Unterstützung der Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, durchgeführt und von der Ethikkommission dieser Universität genehmigt (Ref. ID: IR.AJUMS.REC:1398.84).

Interessenkonflikt

Die Autor*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Anhänge

Verfügbar unter <https://doi.org/10.3205/zma001530>

1. Attachment_1.pdf (144 KB)
Academic and technical words in alphabetical order
- nur auf englisch verfügbar

Literatur

1. Bravo MA, Cervetti A. Teaching vocabulary through text and experience in content areas. In: Graves MF, editor. Essential readings on vocabulary instructions. New York: International Reading Association; 2009. p.141-152.
2. Brezina V, Gablasova D. Is there a core general vocabulary? Introducing the New General Service List. Appl Ling. 2015;36(1):1-22. DOI: 10.1093/applin/amt018

3. Browne C, Culligan B, Phillips J. A New General Service List. 2020. Zugänglich unter/available from: <http://www.newgeneralservicelist.org/>
4. Chen Q, Ge GC. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *Engl Spec Purp.* 2007;26(4):502-514. DOI: 10.1016/j.esp.2007.04.003
5. Chung TM, Nation P. Identifying technical vocabulary. *System.* 2004; 32(2):251-263. DOI: 10.1016/j.system.2003.11.008
6. Chung TM, Nation P. Technical vocabulary in specialised texts. *Read Lang.* 2003;15(2):103.
7. Coxhead A. A new academic word list. *TESOL Quar.* 2000;34(2):213-238. DOI: 10.2307/3587951
8. Durrant P. Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Appl Ling.* 2017;38(2):165-193. DOI: 10.1093/applin/amv011
9. Eldridge J. No, There Isn't an 'Academic Vocabulary,' but...: A Reader Responds to K. Hyland and P. Tse's "Is There an 'Academic Vocabulary'?" *TESOL Quart.* 2008;42(1):109-113. DOI: 10.1002/j.1545-7249.2008.tb00210.x
10. Evans S, Morrison B. Meeting the challenges of English-medium higher education: The first year experience in Hong Kong. *Engl Spec Purp.* 2011;30(3):198-208. DOI: 10.1016/j.esp.2011.01.001
11. Farrell P. Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25. 1990.
12. Gardner D, Davies M. A new academic vocabulary list. *Appl Ling.* 2014;35(3):305-327. DOI: 10.1093/applin/amt015
13. Hsu W. Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *Engl Spec Purp.* 2014;33:54-65. DOI: 10.1016/j.esp.2013.07.001
14. Hyland K, Tse P. Is there an "academic vocabulary"? *TESOL Quart.* 2007;41(2):235-253. DOI: 10.1002/j.1545-7249.2007.tb00058.x
15. Khani R, Tazik K. Towards the development of an academic word list for applied linguistics research articles. *RELJ.* 2013;44(2):209-232. DOI: 10.1177/0033688213488432
16. Krishnamurthy R. Size matters: Creating dictionaries from the world's largest corpus. In: KOTESOL Proceedings 2001. The Learning Environment: The Classroom and Beyond. Proceedings from the 9th Korea TESOL International Conference; 2001 October 13-14; Seoul, Korea. p.169.
17. Laufer B. What percentage of lexis is necessary for comprehension? In: Lauren C, Nordman M, editors. *From humans to thinking machines.* Clevedon: Multilingual Matters; 1989. p.316-323.
18. Laufer B, Nation P. A vocabulary-size test of controlled productive ability. *Lang Test.* 1999;16(1):33-51. DOI: 10.1177/026553229901600103
19. Lei L, Liu D. A new medical academic word list: A corpus-based study with enhanced methodology. *J Engl Acad Purp.* 2016;22:42-53. DOI: 10.1016/j.jeap.2016.01.008
20. Martínez IA, Beck SC, Panza CB. Academic vocabulary in agriculture research articles: A corpus-based study. *Engl Spec Purp.* 2009;28(3):183-198. DOI: 10.1016/j.esp.2009.04.003
21. Mudraya O. Engineering English: A lexical frequency instructional model. *Engl Spec Purp.* 2006;25(2):235-256. DOI: 10.1016/j.esp.2005.05.002
22. Murray N. An academic literacies argument for decentralizing EAP provision. *ELT J.* 2016;70(4):435-443. DOI: 10.1093/elt/ccw030
23. Nagy W, Townsend D. Words as tools: Learning academic vocabulary as language acquisition. *Read Res Quart.* 2012;47(1):91-108. DOI: 10.1002/RRQ.011
24. Nation P. How Good Is Your Vocabulary Program? *ESL Mag.* 2001;4(3):22-24. DOI: 10.1177/105382590102400106
25. North S. Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Stud High Educ.* 2005;30(5):517-533. DOI: 10.1080/03075070500249153
26. O'Flynn JA. An Economics Academic Word List (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *J Acad Lang Learn.* 2019;13(1):A28-A87.
27. Pearson J. *Terms in context.* Amsterdam: John Benjamins; 1998. DOI: 10.1075/scl.1
28. Ryan G. Technical vocabulary acquisition through texts. A corpus and a case study in theology classroom. Saarbrücken: Lambert Academic Publishing; 2012.
29. Shaw P. Science research students' composing processes. *Engl Spec Purp.* 1991;10(3):189-206. DOI: 10.1016/0889-4906(91)90024-Q
30. Veenstra J, Sato Y. Creating an Institution-specific Science and Engineering Academic Word List for University Students. *J Asia TEFL.* 2018;15(1):148. DOI: 10.18823/asiatefl.2018.15.1.10.148
31. Vongpumivitch V, Huang JY, Chang YC. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *Engl Spec Purp.* 2009;28(1):33-41. DOI: 10.1016/j.esp.2008.08.003
32. Wang J, Liang SL, Ge GC. Establishment of a medical academic word list. *Engl Spec Purp.* 2008;27(4):442-458. DOI: 10.1016/j.esp.2008.05.003
33. West MP. A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology. London: Longmans, Green & Company; 1953.
34. Xue G, Nation IS. A university word list. *Lang Learn Comm.* 1984;3(2):215-229.
35. Yang MN. A nursing academic word list. *Engl Spec Purp.* 2015;37:27-38. DOI: 10.1016/j.esp.2014.05.003

Korrespondenzadresse:

Mahmood Maniati
Ahvaz Jundishapur University of Medical Sciences, PO
Box: 159, IR-15794-61357 Ahvaz, Iran
maniatim@yahoo.com

Bitte zitieren als

Tazik K, Maniati M, Afshar MR, Haghighi SB. Developing a new academic wordlist for medical purposes – a viable tool for educators. *GMS J Med Educ.* 2022;39(1):Doc9.
DOI: 10.3205/zma001530, URN: urn:nbn:de:0183-zma0015301

Artikel online frei zugänglich unter
<https://doi.org/10.3205/zma001530>

Eingereicht: 10.01.2021
Überarbeitet: 08.11.2021
Angenommen: 29.11.2021
Veröffentlicht: 15.02.2022

Copyright

©2022 Tazik et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.