

Collocations of high frequency words in nursing research articles and The Academic Collocation List: Similarities and differences

Kaja Mandić & Izabela Dankić

University of Mostar, Bosnia and Herzegovina

Abstract

The main objective of this corpus-based study is to research the most frequent two-word collocations in the corpus of nursing scientific articles and compare this newly assembled list of nursing collocations with the Academic Collocation List (ACL). The nursing scientific articles corpus (NSAC) used in this study comprises 1,119,441 words from 262 articles of 10 high-quality journals from the Medical Library Association list which nursing students can freely access. The focus is on noun-noun and noun-adjective collocations. The selected articles were converted into txt files using the ABBYY Fine Reader. WordSmith Tools 7.0 and TermeX were used for noun and collocation extraction. The newly assembled Nursing Collocation List (NCL) and the ACL were compared using Microsoft Excel 2016. A total of 488 collocations were identified in the NSAC and the NCL contains 234 (47.9%) noun + noun and 254 (52.1%) adjective + noun collocation combinations. The most frequent two-word collocation is *health care* and it appeared 618 times in the NSAC. The ACL (2,469) and the NCL (488) share 123 two-word collocations. Although there are some correspondences between collocations in the two corpora, key nursing collocations with notably higher frequencies are identified in the NSAC (365). Despite the fact that the ACL is the most extensive collocation list across different academic fields and it certainly plays an important role in teaching English as a foreign language, this study suggests that it does not provide key nursing collocations for improvement of nursing collocation competence.

Key words: collocations; nursing, English for academic purposes (EAP); collocation list.

1. Introduction

EAP (English for academic purposes) teachers recognize that courses they teach should meet demanding student needs. It is expected that students are not only to master specific aspects of language, but should also be prepared

to understand and use the language of a particular discipline in order to better communicate, write professional and scientific papers, and publish them. Since English maintains its dominance as the language of science, mastery of disciplinary vocabulary used in authentic texts is therefore very important for EAP learning and teaching.

As far as the nursing academic education is concerned, nursing scientific journals greatly contribute to faster knowledge distribution through dissemination of research findings and development of the nursing profession. Since authentic and relevant texts from these scientific journals prepare nursing students for successful professional and scientific careers, EAP teachers must carefully consider how to help students learn and acquire scientific and professional English of nursing and medical English in general.

As a result of a rigorous selection process and expert review, Ackermann & Chen (2013) created the Academic Collocation List (ACL) derived from the curricular component of the Pearson International Corpus of Academic English (PICAЕ). This list is the most acknowledged collocation list compiled for academic purposes to date comprising over 37 million words of academic written and spoken texts from five major English-speaking countries, i.e. Australia, Canada, New Zealand, UK, and USA. The written part of the corpus includes texts from four fields of study: applied sciences and professions, humanities, social sciences, and natural/formal sciences, and each of them contains materials from seven academic disciplines. Health Sciences including 1,429,679 single words (tokens) are listed within applied sciences and professions. Gledhill (2000) argues that a representative and specialized corpus of medical research articles is needed for teaching languages for specific purposes and the ethnographic aims of genre analysis in general. He made this proposal researching collocations of high frequency words in medical research abstracts and articles.

Budgell et al. (2007) established a need for research of nursing vocabulary and special disciplinary word lists, based on their pilot study of the first issues of 6 nursing journals. In 2015, Yang created a nursing academic word list by analyzing the most frequently used nursing academic vocabulary in nursing research articles. She analyzed 252 nursing articles and 1,006,934 running words (tokens). Pournia (2019) also studied the most frequent academic words in 13 high impact factor English nursing journals from 2012 - 2014 (2852 articles and 8,196,953 tokens) and determined a list of 1081 nursing specific word families.

2. Methods

It could be assumed that mastery of the ACL could lead to nursing collocation competence. Therefore, the main objective of this corpus-based study is to research the most frequent two-word collocations in the corpus of nursing scientific articles and compare this newly assembled list of nursing collocations with the ACL. Since the ACL is a wide-range list of academic collocations it also investigates the need for an independent list of nursing collocations.

This study explores the following research questions:

1. How many collocations from the Nursing Collocation List appear on the Academic Collocation List?
2. Is there a need for an independent nursing collocation list?

Since this study focuses on locating the most frequent two-word collocations used by authors in their scientific research articles, the corpus was carefully collected to serve this purpose. The articles were chosen according to the 2012 evidence-based list of journals for nursing created by the Medical Library Association, and this list is an update and expansion of the Key Nursing Journals List, originally developed in 1986 (Sherwill-Navarro et al., 2014: 105). The list consists of 30 highest listed nursing journals by number of research articles and by percentages of research articles they contain. Only scientific research articles were included in this written corpus because including more than one type of academic prose could affect the results of the study as lexical bundles are register-bound (Bal, 2010: 15).

The nursing scientific articles corpus (NSAC) used in this study comprises 1,119,441 words from 262 articles of 10 high-quality journals from the Medical Library Association list which nursing students could freely access. The size of the corpus was determined according to the target representativeness rather than on size. According to Gray et al. (2017: 1), representativeness in corpus design is crucial, since the goal of most corpus studies is to identify quantitative linguistic patterns in the corpus sample and generalize those findings to a larger linguistic population. The selected journals and the number of research articles used from these journals are presented in Table 1 below.

In order to ensure the representativeness of the nursing scientific articles we used those published in 2017 and 2018 (see Table 1 for a complete list of journals and number of articles). Articles written by the same author were eliminated so the results are not under the influence of the authors' writing style. All the research articles in the corpus were collected in their electronic form and kept at their original length. The selection was done according to Wood's criterion which says that first authors must have native names to the

country concerned and be affiliated with an institution where this language (English) is spoken as the first language (2001: 79).

Table 1: Journals and number of articles used in the study

	Journal	Number of articles
1.	Journal of Clinical Nursing	83
2.	Journal of Advanced Nursing	34
3.	Scandinavian Journal of Caring Science	28
4.	Cancer Nursing	26
5.	International Journal of Nursing Practice	20
6.	Journal of Nursing Management	17
7.	Journal of Nursing Education	11
8.	Journal of Psychiatric and Mental Health Nursing	13
9.	Oncology Nursing Forum	8
10.	Journal of Women's Health	22
Total		262

Only two-word collocations were studied because the majority of collocations in the ACL are two-word noun collocation combinations (1835). The focus was on noun-noun and noun-adjective collocations.

2.1. Data Processing

The selected articles were converted into txt files using the ABBYY Fine Reader software application, after titles, names of authors, literature, tables and charts were deleted manually. The corpus was assembled by an online software program, Corpus Builder v.2.3 as a single combined file. Then, we used the module *WordList* of WordSmith Tools 7.0, an extraction software developed mainly for corpus linguistics to extract the most frequent nouns from the NSAC.

The third and most important software used in the study was TermeX version 1.0. TermeX is a tool for automatic collocation extraction and terminology lexica construction. It is based on statistical measures called association measures (AMs). Fourteen AMs are implemented in TermeX, which, combined with lemmatization, enable users to faster and better construct terminology lexica (Delač, 2009: 2).

Although there are several collocation extraction tools available today, TermeX differs from the rest in that it provides a much wider range of AMs to choose from and it outperforms the majority of tools in terms of processing speed; a corpus of 150 MB can be processed in less than 20 minutes.

In TermeX, extraction is based on statistical measures (AMs) that provide information on how likely is it for an n-gram (a word) to be a collocation. This way, n-grams that are most likely to be collocations become top ranked (Delač et. al., 2009: 150).

The newly assembled Nursing Collocation List and the ACL were compared using Microsoft Excel 2016.

2.2. Criteria for identification and comparison of collocations

In this study, a collocation refers to a group of two words that frequently occur together. A collocation is made up of two parts: the pivot word or the node and the word accompanying it (Shin & Nation 2008: 341).

Shin & Nation's criteria (2008: 341-343) for identifying collocations as well as the ACL's criterion for normed frequency were used with minor adjustments and additions:

1. Each pivot word is a word type. That is, the different word forms 'book' and 'books' were treated as different pivot words and investigated separately, rather than treating 'book' and 'books' as one-word family.
2. The pivot word had to be a noun, a verb, an adjective, or an adverb. We searched only for collocates of content words and in our case content words were only nouns.
3. All the pivot words had to occur in the most frequent 1,000 content words of English according to the spoken word frequency list by Leech, Rayson and Wilson from 2001. This criterion was partly eliminated from our study, since our reference corpus was the Academic Collocation List. But we did focus on high frequency words from our corpus because the working assumption was that learning collocations should strengthen and enrich nursing students' vocabulary of English.
4. According to Shin & Nation each collocation had to occur at least thirty times in ten-million running words. This criterion was completely eliminated. We adopted the normed frequency of 0.2 times per million used during the assembly of the ACL. Ackerman & Chen (2013: 8) raised the threshold during their assembly of the ACL, because it was found that entries with a score less than 4 were mainly noun-preposition combinations and fragments of extended phrases which were not target combinations.
5. Each collocation should not cross an immediate constituent boundary and it should be a complete grammatical unit such as a noun phrase, adjectival phrase or prepositional phrase. For example, 'You at the place' does not meet this criterion because it crosses an immediate

constituent boundary. Single words are also immediate constituents but are of course not collocations.

6. Different senses of collocations with the same form are counted separately.

This study compares two very differently sized corpora. According to Carter & McCarthy (2001) a corpus size depends on what is being investigated i.e. there is no stipulated criterion for corpus size. Flowerdew (2004: 16-19) also argues that there is no ideal size for a corpus - size is dependent on the needs and purposes of the investigation. Specialized corpora are usually smaller in size than generalized ones, and therefore their size and composition make them more manageable for specialized corpus studies. A key aspect of specialized corpora is the comparative nature of many of the investigations, which we do not find to such an extent in the literature on general corpora. Sinclair's statement that "comparison uncovers differences regardless of size" is probably the most suitable argument in this dilemma of comparison of different sized corpora (2001: 13).

We compared collocation patterns in terms of frequency. This way it can be seen whether certain collocations stand out in a specialized nursing corpus when compared to a large general academic corpus.

3. Results

While the list of 30 most frequent nouns, their order of appearance and frequency can be seen in Table 2, the number of collocation combinations including the thirty most frequent nouns is presented in Table 3. We extracted only 30 nouns as defined by the only available option in the demo version of WordSmith Tools 7.0 (See Table 2 for a complete list of the most frequent nouns from the NSAC).

Table 2: The most frequent nouns identified in the NSAC.

NOUN	No. of appearance	No. of frequency
1. care	15	7017
2. patients	24	4901
3. study	26	4515
4. health	29	4069
5. nurse	33	3718
6. nursing	35	3394
7. research	38	2699
8. participants	39	2643
9. data	42	2409
10. cancer	54	1958
11. support	57	1909

12. time	58	1897
13. clinic	61	1828
14. family	62	1826
15. people	69	1675
16. analysis	72	1630
17. staff	73	1610
18. quality	75	1570
19. women	77	1546
20. life	78	1527
21. healthcare	79	1513
22. practice	82	1475
23. experience	83	1450
24. treatment	84	1448
25. group	87	1390
26. information	88	1386
27. work	90	1337
28. report	92	1315
29. physical	93	1309
30. findings	98	1249

The most frequent noun is *care* and it appeared 7,017 times in the overall selection of texts. The most frequent collocation is *health care* which corresponds to the most frequent noun *care* from the same corpus. The noun *care* had the highest number of collocations, 44. Nouns *people*, *women* and *life* had no noun combination collocations (see Table 3 for collocation combinations with the thirty most frequent nouns in NSAC).

Table 3: Number of collocation combinations with the thirty most frequent nouns in the NSAC.

NOUN	No. of collocations	Example	NOUN	No. of collocations	Example
1. care	44	clinical care	16. analysis	7	statistical analysis
2. patient	18	patient safety	17. staff	8	hospital staff
3. study	18	systematic study	18. quality	5	quality appraisal
4. health	22	physical health	19. women	0	-
5. nurse	25	nurse practitioner	20. life	0	-
6. nursing	20	nursing staff	21. health-care	14	healthcare professional
7. research	21	scientific research	22. practice	4	clinical practice
8. participants	1	study participants	23. experience	6	professional experience

9. data	10	qualitative data	24. treatment	4	medical treatment
10. cancer	18	breast cancer	25. group	4	ethnic group
11. support	6	professional support	26. information	3	information sharing
12. time	2	time frame	27. work	12	original work
13. clinic	1	rehabilitation clinic	28. report	2	report findings
14. family	4	family caregivers	29. physical	7	physical environment
15. people	0	-	30. findings	4	key findings

A total of four hundred and eighty-eight (488) collocations were identified in the NSAC. The newly assembled NCL contains 234 (47.9%) noun + noun and 254 (52.1%) adjective + noun collocation combinations (Table 4). In the case of ACL, adjective + noun combinations are also the largest category (71.8%) and noun + noun combinations present only 2.5% of all collocation combinations (see Table 4 for overview of nursing collocations in part of speech combinations). The most frequent two-word collocation is *health care* and it appeared 618 times in the NSAC.

Table 4: Overview of nursing collocations in part of speech combinations.

Type of combination	Number of entries	Percentage	Examples
noun + noun	234	47.9%	<i>heart failure, hospital environment, pain assessment, etc.</i>
adjective + noun	254	52.1%	<i>general practitioner, physical health, recent study, etc.</i>

3.1. Comparison of the ACL and the NCL

The ACL (2,469) and the NCL (488) share 123 two-word collocations (Figure 1). Although there are some correspondences between collocations in the two corpora, key nursing collocations with notably higher frequencies are identified in the NSAC (365).

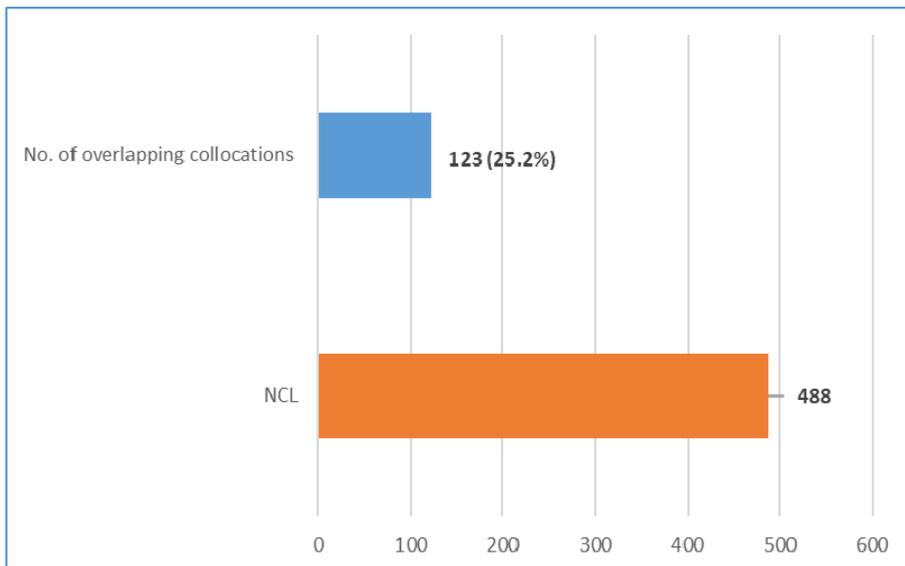


Figure 1: Number of common collocations for the ACL and tNCL.

The most frequently occurring collocations from the Health Sciences field of study that are also listed on both the ACL and NCL can be seen in Table 5.

Table 5: The most frequent collocations from the Health Sciences field of study listed on both lists.

Collocation	Frequency in the NCL	Frequency in the ACL
1. mental health	559	60
2. physical activity	293	5
3. physical health	131	-
4. physical symptom	61	-

Column two shows the value or frequency of these collocations in the NCL and column three the frequency of the same collocations in the ACL. The first most frequent collocation *mental health* from the NCL (frequency value of 559) also appears on the ACL (frequency value of 60). The ACL frequency scores for *physical health* and *physical symptom* are not listed in Table 5, since Ackerman & Chen do not list low frequency entries without the highest expert agreement (Ackerman & Chen, 2013: 15-17).

Mental health is one of four collocations from the field of Health Sciences that appears on both lists. All other collocations that overlap mostly concern academic and scientific writing and are phrases often used in scientific research articles and are generally found across a variety of academic disci-

plines, e.g. *brief review, comparative analysis, current research*, etc. In other words, the majority of collocations found on both lists are not specialized collocations or phrases restricted to the discipline of nursing.

4. Discussion and conclusions

The purpose of this study was to explore two-word noun combination collocations in a selection of scientific articles from nursing journals. After the assembly of the corpus, the objective was to further analyze the extracted collocations with those previously identified by the ACL. What differentiates it from similar EAP studies is that it is particularly focused on academic collocations used in the field of nursing.

Both the ACL and NCL suggest collocations that English language learners are likely to encounter during their academic education. The ACL represents the most important cross disciplinary collocations that can help learners increase their collocational competence and thus their proficiency in academic English. The NCL on the other hand represents collocations that can help nursing students to better acquire English for professional and scientific purposes.

According to the findings from the frequency-based corpus analysis and applied criteria, a list of 488 most frequent nursing collocations was created and compared to the 2,469 collocations from the ACL. It was found that 123 collocations are shared between the two lists - 25.2 % of collocations overlap. However, when collocations from the NCL were compared to those found by Ackerman and Chen (2013), the most frequent collocations did not appear on the ACL. The results suggest that more than three quarters of collocations (74.8%) identified in this study have not been identified in the related literature before. When collocations from the NCL were compared to those identified by the ACL, most of the frequently used collocations did not occur in the PICAE either. Examples of these collocations are: *health care, healthcare professionals, clinical practice, and nursing student*. The most frequent collocations that appear in both the ACL and the NCL are *mental health* and *physical activity*. In other words, collocations from the ACL have a low occurrence rate in the nursing corpus and nursing students also have a low chance of encountering these collocations while reading disciplinary scientific articles.

Noun combinations (adjective + noun) were the dominant type on both lists. The differences between the ACL and NCL may be a result of different text representation in various disciplines. According to Hyland writers need a familiarity with both the clusters which characterize their disciplines and those which are valued in the particular genres of those disciplines (2008: 42).

The ACL can help learners increase their collocational competence and thus their proficiency in academic English (Ackerman & Chen, 2013: 1) but it may be considered most useful for learners who are required to comprehend university academic texts in general. Due to its size, learning and memorizing of the ACL might be more time consuming since it consists of more collocations than the NCL. Despite the fact that the ACL is the most extensive collocation list across different academic fields and it certainly plays an important role in teaching English as a foreign language, it does not provide key nursing collocations. As previously mentioned, a similar corpus-based study was performed by Yang (2015) but it aimed at exploring the most frequently used nursing academic words in a corpus of nursing research articles. The findings of this study also suggest that it is necessary to generate a field specific academic word list for EFL students to strengthen their academic reading and writing proficiency. According to Haswell (in Hyland, 2008: 42) as writers mature, they rely more and more on collocations, and gaining control of a new register requires a sensitivity to preference for certain sequences of words over others. Consequently, learning of those frequent combinations can ensure competent participation in a particular community. Use of more frequent disciplinary phrases can contribute to gaining a communicative competence in a scientific field. The results and conclusions of our study support the argument that more restricted discipline-based collocation and word lists are necessary for individual disciplines.

Despite the fact that the corpus was chosen carefully and under established criteria, there are some limitations. One of them is the fact that the ACL corpus includes both academic written and spoken texts and some previous studies showed astonishing differences between written and spoken corpora. They also show that collocations play a very important role in spoken and written language (Biber 1989; Shin, 2007 in Shin & Nation, 2008: 341). One additional limitation is that the study focuses exclusively on research articles and this may limit representativeness of the nursing scientific articles corpus, although Hyland (2008) points out that research articles are the most prestigious form of academic writing as they are more precise in their aims and structure than other forms of professional academic prose such as textbooks, so they seem to provide the best available model of 'target language' for academic foreign language learners.

Further research could include a mixed-method approach including both qualitative and quantitative research methods. The addition of the qualitative approach might include students' knowledge and competences in using nursing collocations and their feedback. This kind of research could help assemble better lists of academic collocations as well as better discipline specific collocation lists.

References

- Ackermann, Kirsten, Yu-Hua Chen (2013). Developing the academic collocation list (ACL) – a corpus-driven and expert-judged approach. *Journal of English for Academic Purposes* 12.4: 235–247.
- Bal, Betül (2010). *Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars*. PhD Thesis, Georgia State University, Georgia, USA.
- Budgell, Brian, Michiko Miyazaki, Myles O'Brien, Robert Perkins, Yoshiko, Tanaka (2007). Developing a corpus of the nursing literature: A pilot study. *Japan Journal of Nursing Science* 4: 21–25.
- Carter, Ronald, Michael McCarthy (1988). *Vocabulary and Language Teaching*. New York: Longman
- Delač, Davor (2009). *TermeX v1.0*. University of Zagreb, Faculty of Electrical Engineering and Computing.
- Delač, Davor, Zoran Krleža, Bojana Dalbelo Bašić, Jan Šnajder, Frane Šarić, (2009). TermeX: A Tool for Collocation Extraction. In Glebukh, Alexander, ed. *Computational Linguistics and Intelligent Text Processing*. 10th International Conference on Computational Linguistics and Intelligent Text Processing, 149–157.
- Flowerdew, Lynne (2004). The argument for using English specialized corpora to understand academic and professional language. In Connor, Ulla, Thomas A. Upton, eds. *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 11–33.
- Gledhill, Christopher (2000). Collocations in science writing. *Language in Performance Series* 22: 7–20.
- Gray, Bethany, Jesse Egbert, Douglas Biber (2017). Exploring methods for evaluating corpus representativeness. Paper presented at the 9th International Corpus Linguistics Conference, University of Birmingham, July 24–28, 2017.
- Hyland, Ken (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18.1: 41–62.
- Leech, Geoffrey, Paul Rayson, Andrew Wilson (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Longman, London.
- Pournia, Yadollah (2019). A study on the most frequent academic words in high impact factor English nursing journals: A corpus-based study. *Iranian Journal of Nursing and Midwifery Research* 24.1: 11–17.
- Sherwill-Navarro, Pamela, Joy C. Kennedy, Margaret, Allen (2014). Developing an evidence-based list of journals for nursing. *Journal of the Medical Library Association* 102: 105–109.
- Shin, Dongkwang, Paul Nation (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal* 62.4: 339–348.
- Sinclair, John (2001). Preface to small corpus studies and ELT. In Ghadessy, Mohsen, Alex Henry, Robert L. Roseberry, eds. *Small Corpus Studies and ELT. Theory and Practice*. Amsterdam: John Benjamins, 7–15.
- Wood, Alistair (2001). International scientific English: The language of research scientists around the world. Flowerdew, John, Matthew, Peacock, eds. *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press, 71–83.

Yang, Ming-Nuan (2015). A nursing academic word list. *English for Specific Purposes* 37.1: 27–38.

Autors' addresses:

Kaja Mandić
Faculty of Health Studies
University of Mostar
Zrinskog Frankopana 37
88000 Mostar
Bosnia and Herzegovina
e-mail: kaja.mandic@fzs.sum.ba

Izabela Dankić
Faculty of Humanities and Social Sciences
Matice hrvatske b.b.
88000 Mostar
Bosnia and Herzegovina
e-mail: izabela.dankic@ff.sum.ba

Received: May 10, 2020

Accepted for publication: May 29, 2020