UNPACKING THE AUTHENTIC NURSING TERMS: A PILOT STUDY TO A CORPUS-BASED ENGLISH INDONESIAN NURSING DICTIONARY

Komilie Situmorang[1*], Santa Maya Pramusita[2], Hananto[3]

[1-3]Universitas Pelita Harapan

Email Korespondensi: komilie.situmorang@uph.edu

ABSTRACT

Authentic nursing terms help nurses, lecturers, and students give accurate and excellent nursing care. However, there is no existing nursing dictionary that has facilitated specific English-Indonesian nursing terms. This study was a pilot study to develop a nursing corpus to establish bank data and build an authentic nursing dictionary through nursing discourses like journals, clinical notes, scoping, systematic reviews, and commentaries The research utilized corpus linguistics as its working methodology by using purposive sampling to collect 651 articles from seven high-impact nursing journals. The open-access journal PDFs were downloaded and converted into TXT. The TXT files were imported to a concordance software package, Antconc, creating a corpus of 3.613.424 tokens. The frequency analysis resulted in 53.105 words, and cluster analysis was performed to find the word and phrase patterns. The finding revealed a 5000-word pioneer of nursing authentic terms. Later, the words will be translated, and sentence examples will be provided in English and Bahasa Indonesia. These nursing terms will be an excellent source for teaching English to nurses, for the International Nursing Program, nurses in practice, staff, and students in nursing care.

Keywords: Antconc, Corpus Linguistics, Dictionary, Nursing Dictionary, Nursing terms

INTRODUCTION

Effective communication is crucial and paramount in providing nursing care. Nurses need to understand the terminology and its definitions to provide nursing care. As nursing knowledge consists of precise and unambiguous terminology related to critical information related to patient safety, language plays a pivotal role in nursing care practice (Wang et al., 2018). It becomes a solid reason nurses must have a robust nursing dictionary to help them provide accurate nursing care.

Traditionally, nursing terminologies are printed in language dictionaries, which undeniably fails to capture the dynamic nature of nursing languages, which evolve alongside the never-ending advancements in healthcare research, practices, and technology. Printed dictionaries cannot address the nuances intricated in a clinical context. As nursing practices evolve, more new

TAHUN 2025

[MANUJU: MALAHAYATI NURSING JOURNAL, ISSN CETAK: 2655-2728 ISSN ONLINE: 2655-4712, VOLUME 7 NOMOR 1 TAHUN 2025] HAL 237-253

nursing terms have been developed through the nursing practices. It has become one sole reason for calling for an updated digital nursing repository or data bank that meets the nursing advancements.

In nursing education, on the other hand, more nursing programs are being delivered in English, emphasizing the need for nursing terms collection in the form of a dictionary, which will capture this need. The internationalization of higher education has impacted all areas in the university settings in which English has become a precious commodity. In countries where English is spoken as a foreign language, English is used to teach health sciences, nursing education, clinical practices, research, and practices, putting nursing educators and students at a disadvantage as they must be health literate in Nursing English (Lu et al., 2023)

International Nursing Track program, striving for excellence while providing international standard nurses ready to compete in international contexts. The students are prepared to serve in international hospitals and be ready for nurses' mobilization. However, the process has not been easy. Students and lecturers struggle to deliver teaching and learning in English, let alone care for patients. The language barrier was obviously the most challenging issue (Dwipa Arza & Fauzia Rozani Syafei, 2019; Gurun Gunarso & Chen, 2020; Noviana et al., 2019). Undoubtedly, shifting the language of the program has caused challenges. The language of English, in general, is different from nursing English. Fundamentally, nursing terms refer to the professional use of the vocabulary in the nursing area (Davey, 2023). Proficiency in general English differs greatly from using accurate nursing terms when giving

nursing care. There needs to be an existence of a go-to resource that nurses can use to find a solution to an unknown term. Nursing students, especially those in their early years, are new to the field and need a repository that helps them with nursing literacy. As to nursing education, this might mean a data repository is needed to help them increase their health literacy. Thus, nursing education needs the help of an evidence-based nursing dictionary that will accompany their literacy in nursing education.

Vocabulary, in general, is crucial to studying a foreign language and is central to communication (Folse, 2011). Academic vocabulary, on the other hand, accommodates reading, listening, speaking, and writing with sufficient academic lexical terms. Lack of academic vocabulary challenges students' success in academic English in general. (Coxhead, 2000) launched the Academic Word list of 570 essential words to succeed in academic texts to cope with this. This list was a broadened list from the general vocabulary created earlier. The Academic Word list has reaped success in supporting the academic vocabulary in more than 2000 studies. However, these academic nursing terms have not specifically helped unpack the nursing terms.

In nursing, the analysis of such corpora aims to assist health professionals in communicating with their patients (Budgell et al., 2007). The previous study, which analyzed biomedical terms through corpora, found that corpus facilitated health literacy for learners with the terms they encountered in clinical settings (Chung & Nation, 2003; Staples, 2019). To answer these shortcomings, there is an urgency to establish a dedicated nursing dictionary grounded in the principles

of corpus linguistics. Corpus linguistics studies large-scale electronic language data analyzed using computer aid (McEnery et al., 2017; McEnery & Hardie, 2012). To do this, a corpus (plural: corpora), which is a collection of naturally occurring examples of data, must be created first. By creating and analyzing corpora comprising nursing text, such as the current research in nursing practices, academic literature, and clinical notes, researchers can gain invaluable insights into the contextual nuances, lexical patterns, and usage conventions of authentic nursing discourses. Creating a nursing dictionary based on corpus linguistics holds immense potential to help current nursing education, mainly international nursing programs that use English as a medium of instruction. By harnessing the data-driven dictionary, nurse educators, nursing students, nurses, and other health professionals can access a comprehensive repository of specific terminologies relevant to the context, standardized, and evidence-based (Budgell et al., 2007)Moreover, as it is created in electronic form, changes and updates will always be provided along with advancements in nursing care settings to ensure its relevance to current nursing education and diverse clinical settings.

Given this critical importance, this paper is a pilot study establishing the corpus used for electronic language data storage. This study aimed to contribute to the establishment of an English dictionary that offered authentic English nursing vocabulary to the students. Therefore, this study would describe the use of Antconc, a freeware corpus analysis toolkit in collecting the sources, analyzing, and sorting the frequency words. Therefore, this study posed one

research question to be answered: How is the process of unpacking authentic nursing terms using corpus linguistics? To address this question, the study would assess seven highly impacted English journals in one year (2023) and collect 651 articles.

Theoretical Framework
Corpus Linguistics

Corpus (plural corpora) is a large, principled collection of naturally occurring examples of language stored electronically (Kennedy, 2014). The advancement of corpora can be traced back to the late 19th century when computers were first used. Brown Corpus, comprising 1 million words, was created in 1961, but today, corpora are a hundred million words and have contributed massively to language research and teaching. Corpus Linguistics is defined as studying extensive large language data through computer-aided analysis in the form of written or utterances (McEnery et al., 2017). (Kennedy, 2014) emphasizes that corpus linguistics has three excellent and unique features. First, it cannot provide negative evidence, meaning it cannot tell the researchers whether a language is correct. Second, it cannot explain why the finding is as it is because it only presents what the finding is. Finally, the third is that it can provide all possible languages simultaneously, meaning that the language produced by a corpus is not random but planned. However, planned, a large corpus is said not to represent a language as all instances may not be covered.

(Rabadi, 2014) lists some questions that can be answered, including the word and phrase frequency, differences between spoken and written, tense frequency, verb prepositions, specific verb search, modal verbs,

formal vs. informal situations, idiomatic expressions, basic communicative vocabularies, and native speakers vocabularies. These questions contribute to the teaching and learning designs using the corpus. Corpus serves as an empirical method that supports qualitative and quantitative analytical techniques. Therefore, corpora allow us to use the language more accurately. (Kennedy, 2014) expands that the target features of corpus linguistics include phraseology (collocation, lexical bundles, preferred sequences), lexicogrammar, register, English for Specific Purpose), nuances of language, and syllabus design. This very reason highlights why using corpus linguistics is paramount when intending to generate a large repository before building a dictionary. The vocabularies would signify the related terms that are highly used in the discipline, in this context, nursing.

Corpora in Nursing Education

Previous nursing education and healthcare studies have examined various aspects of using corpora communications, clinical settings, research, and nursing education. A study by (Adolphs et al., 2004) looked into the corpora of telephone conversations between the patient and advisers through phone settings, focusing on the strategies used by the health advisors in the interactions and creating a 'convergence coda' approach, which more possibilities in building data and theory. This corpus has then become a resource for health practitioners in clinical settings. (Trinant & Yodkamlue, 2019) established a Sample Corpus of Nursing Articles (SCNRA) containing 1,25 million words and was compiled from 300 research articles from 10 nursing journals. 717 keywords were analyzed to find their part of speech and collocation. The keywords and collocations were used to support teaching and learning in nursing education. In terms of English teaching in the nursing program, a study from Malaysia shows the difficulties experienced by nursing students as they have limited proficiency in English. (Fairus Nor Mohamad & Nor Puteh, 2017) evaluated two English books for Nursing Purposes using Corpus. The results indicated that the books analyzed had 2000 significant words, although not all were presented inside. A similar study on the analysis of textbooks and teaching resources in nursing education was conducted by (Muhammad et al., 2017). The study focused on whether the nursing textbook and nursing journal language differ from the weather report and BNC samplers. The result indicated that the language of the nursing textbooks is highlighted as constrained, unconstrained, and not explicit.

In Indonesia, the use of corpus linguistics to analyze the data in nursing is limited. However, notable studies have looked at and intersected with nursing education or health in general. The study of (Hernina et al., 2023) conducted a corpus linguistic study to establish Indonesian terms of diseases, including 1,2 million tokens from various sources in ten years (2013-2023). It examined the language style of disease names. The finding indicated a particular pattern followed by specific diseases and that linguistic and cultural understanding contributed to terminology uses. Besides, the study of (Fajri et al., 2023) also uses corpus linguistics to analyze the constructions of 'obesity' in Indonesian news media. The study analyzed 1.418 articles that mentioned obesity. This study,

however, focused more on the language stigma and its complexity rather than its implications for nursing education.

From the previous research, it can be concluded that corpora have an extensive impact in generating large and original data. Corpora, generated using corpus linguistics, supports qualitative and quantitative data, providing nuances, including collocations and lexical bundles. Lexicogrammar, and register. Meanwhile, as much as it benefits the gathering of large data, such as making a dictionary, limited studies in Indonesia have utilized corpora to generate authentic nursing terms. Corpus linguistic study has been used for word and term analyses, as provided in the previous paragraphs. Against this previous study, this current study calls for the establishment of nursing terms, originally taken from academic works surrounding nursing, to be later translated into Bahasa Indonesia to help with teaching and learning in the nursing department. With this said, to this day, very limited corpus linguistics studies have been used to comprise authentic nursing education materials and dictionaries in Indonesia. Thus, this study aimed to assess the practicality of developing and analyzing a corpus of the English-language nursing literature to identify the most frequently used words and, hence, words that perhaps ought to be the focus of nursing English instruction and testing (nursing dictionary).

RESEARCH METHOD

The method used was corpus linguistics to create a resource based on proof that adapts to changing nursing needs. By revealing complex language patterns, corpus linguistics can help nursing students in various clinical situations. By its nature, the study used Corpus Linguistics as its methodology (McEnery et al., 2017). However, it also utilized descriptive research and development (R&D) as its working methodology. The R&D model used in this research follows Luther's six-stage model: concept, design, material, collection, assembly, testing, and distribution (Septinaningrum et al., 2021; Susilo et al., 2021).

In this early R&D, however, the steps were yet in the first to the fourth stages, as the fifth and sixth stages would be done through the application PHP programming after securing the funding. The corpus linguistic was used in the *concept, design, material, and collection* by gathering large, principled electronic data from high-impacted English-language Scopus Q1 nursing journals. As the data collection was purposeful, purposive sampling was used (Creswell & Creswell, 2018). Using purposive sampling, this study searched through seven nursing journals, including *Journals of Clinical Nursing, Journal of Advanced Nursing, Primary Health Care Research and Development, International Journal of Community Based Nursing and Midwifery, Nurse Education Today, Journal of Emergency Nursing, and International Emergency Nursing Journal*. To avoid biases, these are chosen based on some of the courses offered in the nursing department. More journals will be included later based on the courses. The table is as follows:

Table 1. The number of open-access articles downloaded

Journals	No of Articles
Journals of Clinical Nursing	240
Journal of Advanced Nursing	183
Primary Health Care Research and Development	68
International Journal of Community-Based Nursing and Midwife	60
Journal of Emergency Nursing	18
Nurse Education Today	62
International Emergency Nursing Journal.	20

The search yielded 651 articles comprising empirical research articles, editorials, commentaries, reviews, systematic reviews, and scoping reviews, which were converted and saved in simple text format (TXT) (Access: bit.ly/ArticleMetadata). 11 students were employed as research assistants to help convert the data. Each research assistant worked 15 hours to convert the data, averaging 59-60 articles each. The research assistants were given the share-point link so they would have access to the downloaded articles. They worked independently through the process and uploaded the converted file to the folder link given to them.

The TXT files were separately coded according to their sections within the paper (, <introduction> <literature review> <methodology> <finding> <discussion>, and <conclusion>). These are intended for future research opportunities. These Txt articles were edited and placed in a single folder to form the pilot corpus. The metadata, including the title, author(s), year, and abstract, was registered in an Excel file as future references to the data used in the corpus. However, it is worth noting that only articles with open access were retrieved as the researchers needed access to the full paper of each journal. This also highlighted the limitations of the study.

The Txt files of each journal were put into a folder and then uploaded to a software package to create a corpus. The corpus was analyzed using the concordance software package Antconc (L Anthony, 2023). The application counted the number frequency used of each different word and resulted in a worksheet listing words and frequency they appeared in the corpus. The word count, the 'token' of the working corpus, was 3.613.424 English words. Approximately 53.105 different words were listed and ranked according to frequency, which had the potential to be used as entries to the prospective nursing dictionary. The function words were separated from the content words. Meanwhile, later in the development of the dictionary, the content words would be seen through Cluster analysis to expand the vocabulary search within the corpus for their word pattern and group words used. Finally, Keywords in context (KWIC) helped define their uses when designing the dictionary. These stages would result in a list of words that would be entries for the nursing dictionaries. The instances of this process are provided in the result section.

Finding and Discussion
Finding

This section presents a comprehensive step-by-step analysis of each stage of corpus linguistics,

TAHUN 2025

[MANUJU: MALAHAYATI NURSING JOURNAL, ISSN CETAK: 2655-2728 ISSN ONLINE: 2655-4712, VOLUME 7 NOMOR 1 TAHUN 2025] HAL 237-253

including corpus creation, reading the tokens, frequency analysis, cluster analysis, and KWIC analysis. In the first step, the 651 articles were converted to TXT. A sample of the TXT file can be seen in picture 3.1.
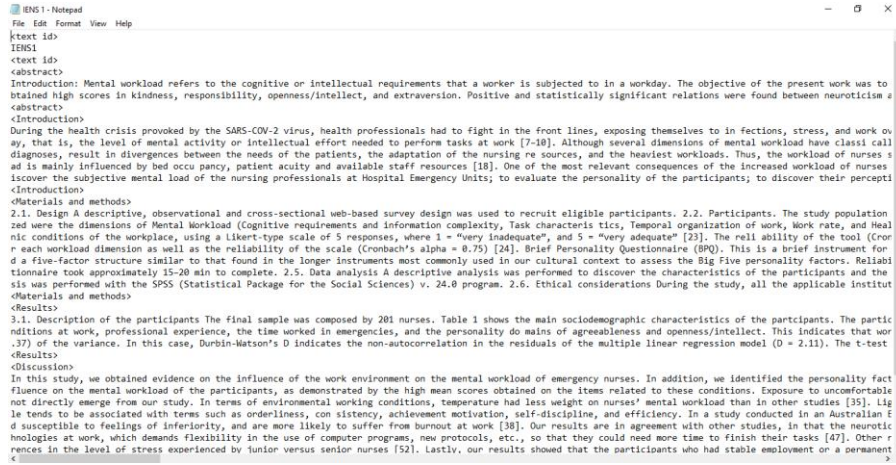
Picture 1. Sample of TXT file

The above picture depicts the sample TXT file required before it was imported to concordance software. All the articles were manually convert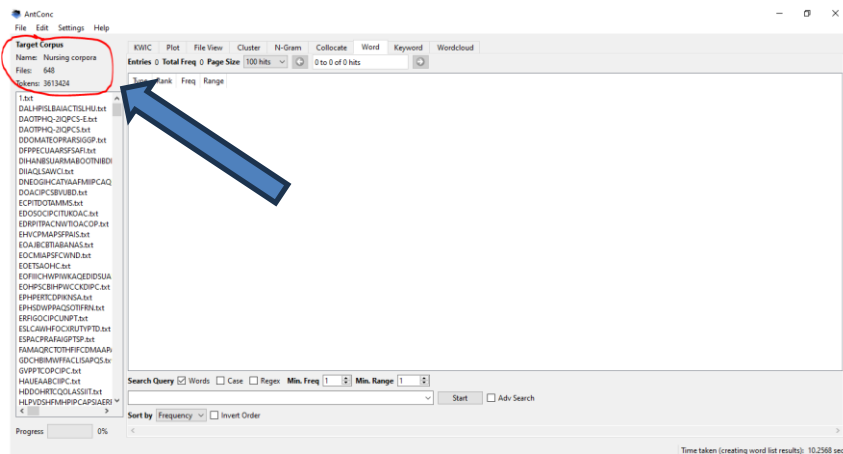ed from PDF to TXT to ensure all the information was copied and pasted as it was. The results were put in a folder to enable the import. The PDF conversion process to TXT took two weeks, and research assistants were employed.
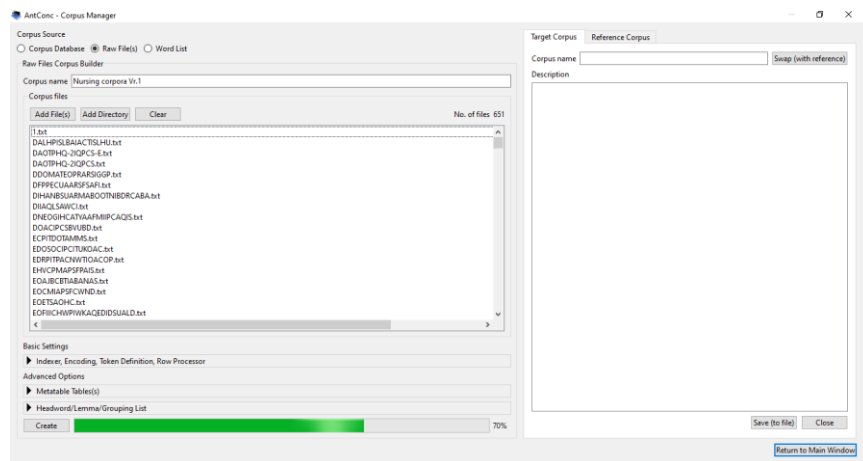
Picture 2. Corpus creation with Antconc 4.1.4

The above picture shows the process of importing the TXT file into the Concordance software 'Antconc'. The process went through opening the Antconc, go to File, choose open Corpus Manager, click Raw file(s), type Corpus name, choose Add Files, go to the folder of TXT you have provided earlier, CTRL +A all the files and click open, click Create to start generating corpus. It took a few seconds to a minute to import all the TXT into the corpus, as shown above

TAHUN 2025

[MANUJU: MALAHAYATI NURSING JOURNAL, ISSN CETAK: 2655-2728 ISSN ONLINE: 2655-4712, VOLUME 7 NOMOR 1 TAHUN 2025] HAL 237-253

Picture 3. Tokens generated in Nursing Corpus.

The above picture shows the newly generated corpus. The import resulted in a new corpus that will be used to search for the authentic nursing terms from the high impacted nursing journals corpus was named after nursing corpus and it resulted in 3.613.424.
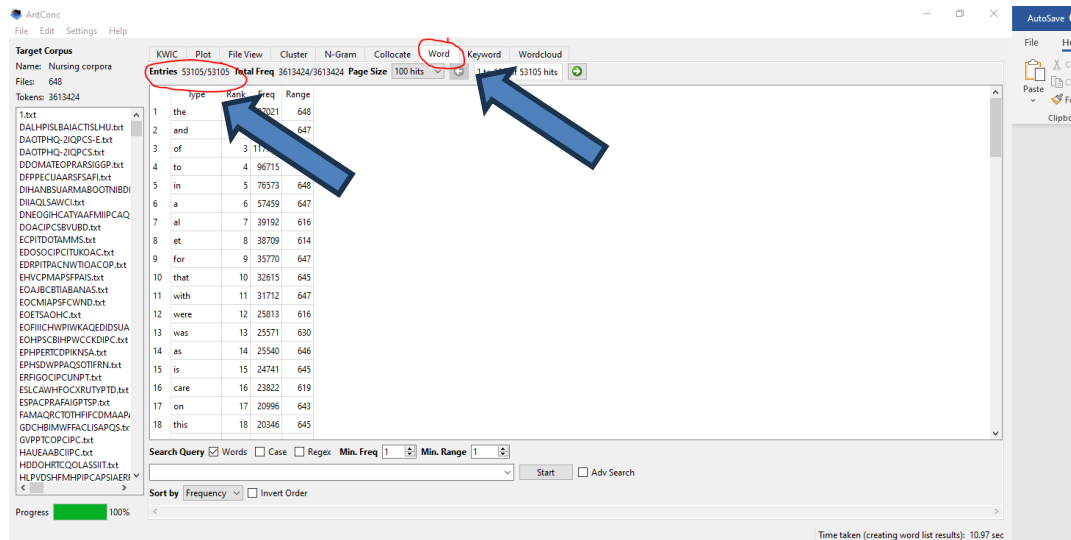
Picture 4. Frequency analysis

The above picture shows the result of the frequency analysis. Frequency analysis was done to search for the word with highest frequency, repeatedly appearing, showing its importance to be learned. To do this, click on the word function and the frequency would appear. The frequency analysis resulted in 53.105 words which as expected, the information about the 'functional words' such as "the,' 'and,' 'of,' 'to' etc., appeared to be the most common words. Ignoring the 'functional words, the content words were separated by downloading the frequency words into the Excel file to sort them. The content words with bigger frequency were sorted

instead of using the alphabetic order.

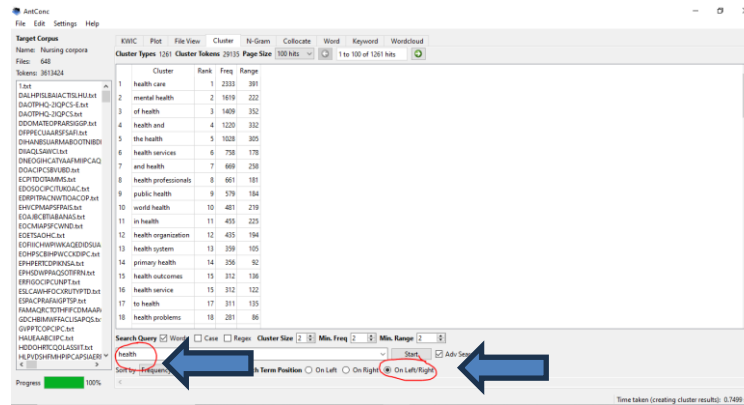**Table 2. Sample content words generated from the frequency words.**

Content Words	Frequency
Health	13345
Nursing	9465
Patients	7276
Clinical	4591
Communication	2439
Intervention	2168
Community	2148
Hospital	2098
Violence	2087
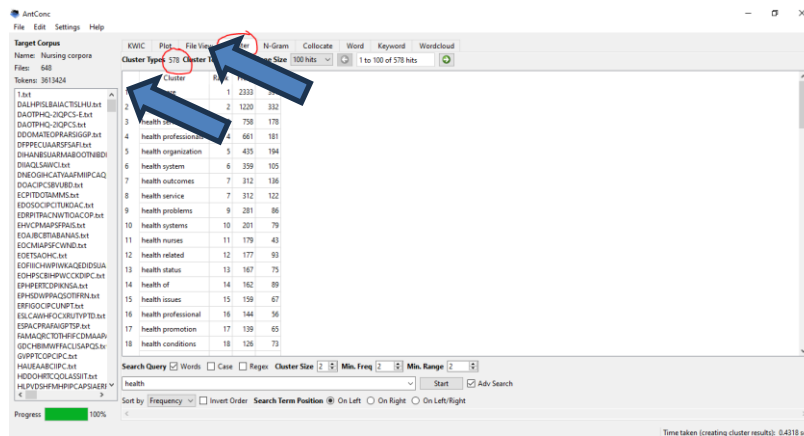
Picture 5. Cluster Analysis

The above picture shows the Cluster Analysis done to each content word. After the completion of the frequency analysis, the content words were listed and brought back into the corpus to search for their clusters in the corpus. Clusters indicate their patterns as words and phrases used within the corpus. The cluster analysis looked through the pairs within the left and the right patterns to guarantee the authenticity and accuracy of the words. The process of finding the word pattern through cluster analysis is by taking the content words into the function. In the above picture, the word 'health' is put into the cluster analysis and searched thoroughly through *left* patterns. 578 cluster words were found from the *left* search position.

TAHUN 2025

[MANUJU: MALAHAYATI NURSING JOURNAL, ISSN CETAK: 2655-2728 ISSN ONLINE: 2655-4712, VOLUME 7 NOMOR 1 TAHUN 2025] HAL 237-253

Picture 6. Word cluster on right/left

The above picture shows the instance of the word 'health' searched on *Left/Right*. A total of 1261 patterns were generated, yet the words went through screening following inclusion or exclusion. The words with preposition patterns like 'of health, the health, and health" were excluded. Meanwhile, the word pattern like "health care, mental health, health services, public health, health professionals, health system, primary health, health service, and health problems were included in the nursing terms words' pioneers.
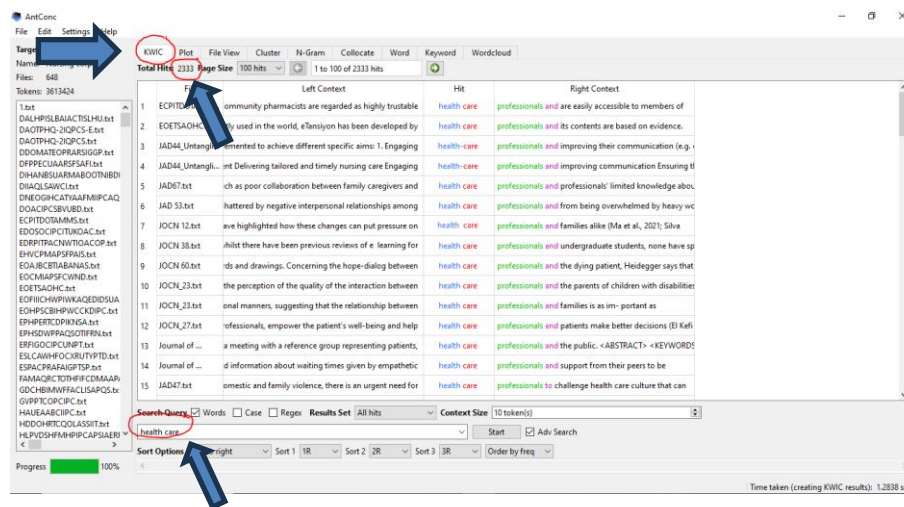
Picture 7. KWIC function analysis

The above picture shows the instance of the KWIC analysis, which will be done later if the nursing terms are to be translated into Bahasa Indonesia and provided its sample sentence. The sentences would emerge by clicking into the KWIC and then entering the words to be searched. The word 'health care' was inserted in the sample above, and 2,333 sample sentences appeared in the context. These sentences will later help define the words, and later, one of them will be picked as an authentic example sentence. The process done repeatedly to search for the content words through a duration of 1 year resulting in 5000 authentic nursing words. The words, however, will not be included here to avoid data breaches.

DISCUSSION

This study sought to unpack the nursing dictionary through corpus linguistics. By mining the open-access high-impacted journal articles of a year length (2023), the study collected 651 articles with 3.613.424 tokens and approximately 53.105 different words with the highest frequencies. To help the process, the study utilized Antconc as the corpus linguistic software package to find the high-frequency words, clusters for word patterns, and the keyword or in context (KWIC) to help with the definitions.

The frequency analysis has resulted in valuable insights into the occurrence of the functional and content words yielded from the corpus. The functional words, such as "the, in, on, for, and too, "were disregarded in the analysis, while content words, such as "health, care, nurses, patient," underwent cluster analysis to find their underlying patterns. These findings represent a crucial step in the preliminary data collection for dictionary entries, offering a structured understanding of how words are used within a specific context. The researchers separate content words and subject them to cluster analysis to identify the recurring patterns or themes in their usage. This cluster analysis is done as it helps uncover the authentic nuances and subtleties in the language used in the corpus, which may not occur through the simple frequency counts alone. Contextual relevance and semantic relationships were revealed in a clearer picture after seeing the words with similar contexts grouped through cluster analysis.

Moreover, the KWIC function helps generate the correct definitions of the content words. Examining their use in their respective contexts and bringing them to the dictionary as authentic material, nuances, and appropriate meaning capture the message and use. KWIC analysis ensures that the dictionary entries are comprehensive and reflective as they are words used in real-world contexts. Overall, integrating these triple analyses, frequency, cluster, and KWIC analysis, represents a robust approach to understanding and documenting the authentic word pattern within the corpus. The preliminary findings through the R&D and corpus linguistics analysis lay the groundwork for developing the nursing dictionary, providing accuracy and relevance to the terms used in nursing practices. This breakthrough will be refined and validated to contribute to advancing the nursing dictionary R&D and developing nursing education by providing comprehensive language resources.

This study has shown how the corpus linguistic approach to establish the 'evidence-based nursing dictionary' provides efficiency in terms of cost and time to seek the original languages used in the nursing journals. As the previous study (Kennedy, 2014) suggested, this approach has expanded the nuances of the nursing terms used in the corpus created. Through this approach, authentic word/phrase patterns were found in the nursing journals, which were the reflection from the field, the clinical practices, and the evidence from the nursing practices. Through this corpus linguistic approach, the dictionary will support nursing education where the language provided aligns with the actual vocabulary in the field and enhances comprehension and the communication skills of the lecturers, students, and those already in the nursing profession. (Lu

TAHUN 2025

[MANUJU: MALAHAYATI NURSING JOURNAL, ISSN CETAK: 2655-2728 ISSN ONLINE: 2655-4712, VOLUME 7 NOMOR 1 TAHUN 2025] HAL 237-253

et al., 2023; Panda et al., 2021; Wang et al., 2018)The previous study has much to say concerning the challenges of nurses' English. This study would provide a resource to help cope with the challenges as it gives access to the most used terms of nursing.

Corpus linguistics has also been used to analyze the prosodic features of nurse-patient interactions by comparing the discourses of U.S. and internationally educated nurses (Staples, 2019; Wagner et al., 2015). The finding indicates a difference in pronunciation, where international nursing students' pronunciation was more pragmatically appropriate than that of US nurses. The research later became the basis for integrating a pronunciation course.

As given in the background of this study, the initial purpose was to support nursing students and lecturers who were put into the international nursing program and were struggling with the vocabulary in nursing. In fact, vocabulary studies have mushroomed as vocabulary is vital in language learning (Folse, 2011). Vocabulary enables students to operate in the target language to deliver messages and to accept messages at the same time. However, it is worth noting that vocabulary is different in different disciplines.

Vocabularies are vital in learning a language (Folse, 2011). They accommodate comprehension and are the body of the words used in the target language. In an academic context, academic vocabulary contains words in academic texts. These are words to succeed in academic reading, writing, speaking, and listening. Academic vocabulary enables learners to make academic analyses, inferences, syntheses, and conclusions (Schmitt, 2008). It is a

vital skill because learners must understand the content vocabulary to understand the concept, they are learning academically (Choo et al., 2017). In 1972, a General service List was created as the first Academic Word List (AWL). This list was extended to 570-word families and became the existing AWL today. The AWL has been said to revolutionize the learning of English for academic purposes (Yang, 2015)However, this list was not explicitly related to nursing or any other specific discipline; it was only emphasized as being very important for university students to master.

English vocabulary, in general, is significantly different from those in nursing. Some vocabulary specifically belongs to nursing. Vocabulary in nursing is the English for Specific Purposes (ESP) (Pradana et al., 2022). Fundamentally, nursing terms refer to the professional use of the vocabulary in the nursing area. Besides, as nursing science is authentically written in English, access to it will grow familiarity with the authentic material. In the process, students will embrace and comfortably experience them while learning and using them with their patients. Knowing the standard terminology will affect the quality of creating nursing documentation (Yuwanto et al., 2023). A study on the academic nursing word list was conducted in 2015. With a similar approach to this running research, (Yang, 2015) created a corpus of 1,006,934 words from 252 nursing journals. The study resulted in 676-word families, which was then called a nursing academic word list. The study aimed to create a pool of word lists to help people who wanted to publish papers in nursing journals. However, the wordlist is exclusive to English and is not specifically designed for Indonesian nurses.

This study looks at how the lack of a complete nursing dictionary that keeps up with changing nursing terms affects communication in clinical situations. The fact that English is becoming increasingly important in nursing education makes it harder for traditional dictionaries to keep up with the changing language of nursing. The novelty of this study lies in its nature, as English for a specific purpose, that is, the nursing discipline. To trace back, there have been studies generating the creation of the General Word List, which was created to support English language learning in general (West, 1972). This list, however, has been irrelevant as it was created many decades ago, and its purpose was for general English operations. In 2000, an academic word list (AWL) was also composed, comprising 570-word families, becoming a greatly referenced word list today (Coxhead, 2000). The AWL becomes the source for people to learn academic skills and understand academic texts. (Coxhead, 2000) emphasizes that these words are the standard of the words that a university student must have and thus are most studied. The sub-list of these AWLs includes arts, commerce, law, and biology. However, none of the AWLs have a specified section for the nursing field.

On the other hand, the study of nursing is specifically tied and connected to specific terms that are unique only to the nursing field. Fundamentally, nursing terms refer to the professional use of the vocabulary in the nursing area (Pradana et al., 2022). The terms are needed to understand the nursing texts in teaching and learning, nursing journals, the current nursing literature, delivering care, writing clinical notes, and operating in the nursing context. These are the reasons why the general service list and the AWLs are not relevant to the context of nursing. The only relevant study that has touched the core of the problem of this research is the study of nursing academic word lists, which was done in 2015 (Yang, 2015). The study had a similar approach to the working research in which a corpus was created to generate the nursing academic word list. The study resulted in 676-word families; however, it had a different purpose to this study. The nursing academic word list (NAWL) study aimed to help those who wanted to be able to understand the nursing texts and those who wanted to publish in nursing journals. Using the same corpus linguistics approach, however, the current study emphasizes helping nursing students and staff struggling with their nursing English. The previous study of NAWL only investigated the vocabulary that would support the understanding and the academic purpose of the NAWL, which highlights the difference between the two studies. However, it is safe to say that one of this study's goals also intersects with the previous study.

Apparently, there have also been dictionaries of healthcare which cover the term of health in general, like Baillere's Dictionary for Nurses and Health Care Workers for the 27th Edition (Taylor, 2020) and Mosby's Dictionary of Medicine, Nursing & Health Professions (Mosby, 2021). However, these dictionaries are not specifically designed for nurses. More of the emphasis of these dictionaries is to present English technical translation of health terms in general. Another specific nursing dictionary, Nurse's Pocket Guide: Diagnoses, Prioritized Interventions, and Rationales Field (Doenges, 2004), is a significant

resource for nursing terms to be used in operating a hospital as a nurse. Yet, the dictionary is written fully in English and has no Indonesian translation. This is one of the fundamental gaps this study tries to fill in, which is to present a freshly created corpus that resulted in words currently used in nursing discourses today, to be later given the Indonesian meaning to support the teaching and learning in international nursing programs in Indonesia.

The current study collects 5000 words, which will be grouped into the nursing word families. However, the contribution has at least been seen in four ways. First, the nursing academic list with Indonesian meaning is created to support the teaching and learning in the international nursing faculty. There will be a list of words with their definition in Bahasa Indonesia to be read. This will greatly help, as the pioneer words mined from the corpus will be translated into Indonesian. Therefore, it will be a resource for teachers and students when confronting an unknown specific nursing term. Second, this study would be a great resource for creating a nursing curriculum. The nursing lecturers in the international program can refer to the word list to create evidence-based teaching material. The material will contain authentic and updated vocabularies in the field. The nursing word list will also help create the testing materials. Therefore, this word list would help them to operate on basic nursing tasks. Third, this contribution intersects with the previous study of the NAWL (Yang, 2015), that is, to become a resource of a pool of words that allows nursing lecturers and students to publish in nursing journals. We consider it important as publishing in nursing journals has been listed as the

graduation requirement by the Indonesian government. Fourth, the final form of the current nursing terms word list is projected to be the future nursing English-Indonesian dictionary, which will be available to a larger nursing audience. The final refined dictionary would be designed and published within a web-based dictionary or Android application, which may occur in a few years.

With all these said, this study is still in its early descriptive research and development phase. The next feasible step after this study would be defining the acquired 5000 words to Bahasa Indonesia and grouping them into the nursing word families to be used as an authentic resource of nursing terms in teaching and learning, creating material and tests, and writing academic journals at the international nursing program. At the same time, this study would continue its phase of designing web-based nursing dictionaries to provide resources to the larger nursing audience in Indonesia.

However, as this study emphasizes, the limitation lies in the nature of the research. As the working title indicates, "a pilot study," this research is only in its descriptive phase. The current research encompasses the concept, design, material, collection, and collection phases, which have yet to go through the assembly, testing, and distribution phases, as mentioned in the methodology section. Despite these limitations, this study has marked a crucial trajectory in creating the corpus to collect the authentic nursing word list to contribute to teaching and learning in nursing education and later contribute to formulating the English-Indonesian nursing dictionary. Thus, future research will translate the 5.000 words and

provide the sample sentence using the help of the KWIC function.

CONCLUSION

This study piloted corpus linguistics to unravel the complexity of English nursing terms. A corpus of nursing was created through corpus linguistics and descriptive research and development methodology, taken from seven high-impact nursing journals. 651 articles were imported into the corpus software package Antconc, resulting in 3.613.424 tokens. Through several analyses of frequency, cluster, and keywords in context analysis, 5000-word pioneers were collected. This study has significantly facilitated the development and analysis of the most frequent words using corpus linguistics because the authentic vocabularies generated are the words most used in research publications.

These words contribute to the struggle of the teaching and learning process in the international nursing program in the faculty of nursing Universitas Pelita Harapan, highlighting the presence of the pool of word lists that will be useful for material design, tests, and future dictionaries. It unpacked the importance of having a nursing dictionary that provided authentic words from the field of nursing education to support the students, lecturers, and nurses in practice, as no dictionaries of nursing in Indonesia have attempted to do so. The implication of this study extends to creating an authentic nursing word list with Indonesian meanings that will help nursing education, nursing practices, and nursing academics.

REFERENCES

Adolphs, S. S., Brown, B., Carter, R., Crawford, P., & Sahota, O. (2004). Applying Corpus Linguistics In A Health Care Context. *Journal Of Applied Linguistics*, *1*(1), 9–28. Https://Doi.Org/10.1558/Japl.1.1.9.55871

Anthony, L. (2023). *Antconc (4.2.4)*. Waseda University.

Budgell, B., Miyazaki, M., O'brien, M., Perkins, R., & Tanaka, Y. (2007). Developing A Corpus Of The Nursing Literature: A Pilot Study. *Japan Journal Of Nursing Science*, *4*(1), 21–25. Https://Doi.Org/10.1111/J.1742-7924.2007.00071.X

Choo, L. B., Lin, D. T. A., Singh, M. K. M., & Ganapathy, M. (2017). The Significance Of The Academic Word List Among Esl Tertiary Students In A Malaysian Public University. *3l: Language, Linguistics, Literature*, *23*(4), 56–65. Https://Doi.Org/10.17576/3l-2017-2304-05

Chung, T., & Nation, P. (2003). Technical Vocabulary In Specialised Texts. *Reading In A Foreign Language*, *15*.

Coxhead, A. (2000). A New Academic Word List. *Tesol Quarterly*, *34*(2), 213–238. Https://Doi.Org/Https://Doi.Org/10.2307/3587951

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative And*

Mixed Methods Approaches (5th Ed.). Sage.

Davey, A. K. (2023). Internationalisation Of The Curriculum In Health Programs. *Bmc Medical Education*, *23*(1), 285. Https://Doi.Org/10.1186/S12 909-023-04271-8

Folse, K. (2011). Applying L2 Lexical Research Findings In Esl Teaching. *Tesol Quarterly*, *45*. Https://Doi.Org/10.5054/Tq. 2010.254529

Gurun Gunarso, A., & Chen, S.-C. (2020). *The Challenges And Needs Of Nurses' English Communication: A Case Study In Taipei City Hospital In Taiwan*.

Hernina, H., Karlina, Y., & Puspitasari, D. A. (2023). The Indonesian Terms Of Disease Names: A Corpus Linguistic Study. *Okara: Jurnal Bahasa Dan Sastra*, *17*(1), 14–31. Https://Doi.Org/10.19105/Oj bs.V17i1.8571

Kennedy, G. (2014). An Introduction To Corpus Linguistics. *An Introduction To Corpus Linguistics*, 1–315. Https://Doi.Org/10.4324/978 1315843674

Lu, L., Jin, Y., Liao, R., Chuang, Y. C., & Tung, T. H. (2023). English Training Requirements And Associated Factors For Non-Native English-Speaking Nurses: A Critical Gap Analysis Based On The Importance-Performance Method. *Heliyon*, *9*(6), E16481. Https://Doi.Org/10.1016/J.H eliyon.2023.E16481

Mcenery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory And Practice. In *Corpus Linguistics: Method, Theory And Practice*.

Https://Doi.Org/10.1017/Cbo 9780511981395

Mcenery, T., Love, R., & Brezina, V. (2017). Introduction. *International Journal Of Corpus Linguistics*, *22*(3), 311–318. Https://Doi.Org/10.1075/Ijcl. 22.3.01mce

Mosby. (2021). *Mosby's Dictionary Of Medicine, Nursing & Health Profession* (11th Ed.). Elsevier.

Muhammad, M. M., Hamzah, S. G., Abdullah, S. K. Bin, & Jack, C. S. (2017). A Study Of Closure In A Nursing Textbooks And Journals: A Corpus Based Study. *International Journal Of Advanced And Applied Sciences*, *4*(2), 96–105. Https://Doi.Org/10.21833/Ija as.2017.02.017

Noviana, I., Nahariani, P., & Rosmaharani, S. (2019). The Problems Faced In Learning English For Nursing Of Students In Stikes Pemkab Jombang. *Kne Life Sciences*, *2019*, 535–541. Https://Doi.Org/10.18502/Kls .V4i13.5303

Panda, S., Dash, M., John, J., Rath, K., Debata, A., Swain, D., Mohanty, K., & Eustace-Cook, J. (2021). Challenges Faced By Student Nurses And Midwives In Clinical Learning Environment – A Systematic Review And Meta-Synthesis. *Nurse Education Today*, *101*(November 2020), 104875. Https://Doi.Org/10.1016/J.N edt.2021.104875

Pradana, A., Yunita, W., & Diani, I. (2022). What Do Nursing Students Need In Learning English? *Joall (Journal Of Applied Linguistics And Literature)*, *7*(2), 321–344. Https://Doi.Org/10.33369/Jo all.V7i2.14819

Rabadi, R. (2014). The Use Of Corpora In Linguistics. *5*, 147–166. Https://Doi.Org/10.12816/0015404

Schmitt, N. (2008). Review Article: Instructed Second Language Vocabulary Learning. *Language Teaching Research*, *12*(3), 329–363. Https://Doi.Org/10.1177/1362168808089921

Susilo, A., Hardyanto, W., Martuti, N., & Purwinarko, A. (2021). Mobile Learning Development Using Augmented Reality As A Biology Learning Media. *Journal Of Physics: Conference Series*, *1918*, 42013. Https://Doi.Org/10.1088/1742-6596/1918/4/042013

Taylor, J. (2020). *Bailliere's Dictionary For Nurses And Health Care Workers* (27th Ed.). Elsevier.

Trinant, K., & Yodkamlue, B. (2019). Lexical Collocations In A Sample Corpus Of Nursing Research Articles (Scnra). *Suranaree Journal Of Social Science*, *13*, 45–72. Https://Doi.Org/10.55766/Obxc3400

Wagner, L. M., Brush, B. L., Castle, N. G., Eaton, M., & Capezuti, E. (2015). Examining Differences In Nurses' Language, Accent, And Comprehensibility In Nursing Home Settings Based On Birth Origin And Country Of Education. *Geriatric Nursing*, *36*(1), 47–51. Https://Doi.Org/Https://Doi.Org/10.1016/J.Gerinurse.2014.10.012

Wang, Y. Y., Wan, Q. Q., Lin, F., Zhou, W. J., & Shang, S. M. (2018). Interventions To Improve Communication Between Nurses And Physicians In The Intensive Care Unit: An Integrative Literature Review. *International Journal Of Nursing Sciences*, *5*(1), 81–88. Https://Doi.Org/10.1016/J.Ijnss.2017.09.007

West, M. (1972). *A General Service List Of English Words*. Longma: Green And Co.

Yang, M.-N. (2015). A Nursing Academic Word List. *English For Specific Purposes*, *37*, 27–38. Https://Doi.Org/Https://Doi.Org/10.1016/J.Esp.2014.05.003

Yuwanto, M. A., Eliya Astutik, E., & Eko Prasetyo, R. (2023). Application Of Nursing Terminology Based On The Indonesian Nursing Standards (Diagnosis, Outcome, And Intervention) On The Quality Of Filling Out Documentation. *Jurnal Keperawatan*, *14*(02), 106–111. Https://Doi.Org/10.22219/Jk.V14i02.24106