**INDONESIAN JOURNAL OF APPLIED LINGUISTICS**
Vol. 14 No. 3, January 2025, pp. 553-568

Available online at:
https://ejournal.upi.edu/index.php/IJAL/article/view/67603

https://doi.org/10.17509/ijal.v14i3.67603

# Steps for developing the English Corpus of Public Administration (ECOPA) for public administration students: A qualitative corpus approach

**Misnawati[1*], Haryanto Atmowardoyo[2], Iskandar Sulaiman[3], Saidna Zulfiqar bin Tahir[3], and Yusriadi**

[1]*Faculty of Social Sciences and Humanities, Universitas Cahaya Prima, Jl. Urip Sumoharjo, Bone, Sulawesi Selatan, Indonesia*

[2,3]*Faculty of Language and Literature, Universitas Negeri Makassar, Jl. Mallengkeri Raya, Parang Tambung, Makassar, Sulawesi Selatan, Indonesia*

[4]*Faculty of Language and Literature, Universitas Iqra Buru, JL. Universitas, Namlea, Buru, Maluku, Indonesia*

[5]*Postgraduate Program in Administrative Sciences, Universitas Cahaya Prima, Bone, Indonesia, Jl. Urip Sumoharjo, Bone, Sulawesi Selatan, Indonesia*

**ABSTRACT**

Corpus linguistics has opened up exciting possibilities for enriching classroom learning, yet only some corpus-based tools exist, specifically for didactic purposes. This is particularly true in public administration, where a lack of dedicated corpora presents a significant research opportunity. This study addresses this gap by developing the English Corpus of Public Administration (ECOPA), a novel corpus-based didactic tool designed to enhance the learning and teaching of public administration vocabulary. This study employed a qualitative approach grounded in corpus linguistics principles, utilizing the corpus analysis steps outlined by Toriida (2016) and a semantic scale validation process adapted from Dang (2020). ECOPA was meticulously compiled from 561 written text references representing diverse genres relevant to public administration, including narrative, general, academic, formal, and informal texts. Analysis of this corpus yielded a comprehensive list of 6.283 online words pertinent to the field. ECOPA is a valuable resource for public administration learners, as it provides exposure to authentic language use and facilitates vocabulary acquisition. Lecturers can utilize ECOPA to design engaging language-focused activities and assessments, while scholars can leverage the corpus for further research into the linguistic landscape of public administration. This study underscores the profound impact of corpus development on enriching linguistic resources within specialized fields.

**Keywords:** Corpus linguistics; English Corpus of Public Administration (ECOPA); qualitative corpus approach; semantic scale; specialized corpus

| **Received:** | **Revised:** |
|---|---|
| 1 March 2024 | 21 August 2024 |
| **Accepted:** | **Published:** |
| 19 December 2024 | 31 January 2025 |

**How to cite (in APA style):**

Misnawati, M., Atmowardoyo, H., Sulaiman, I., Tahir, S. Z. b., & Yusriadi, Y. (2025). Steps for developing the English Corpus of Public Administration (ECOPA) for public administration students: A qualitative corpus approach. *Indonesian Journal of Applied Linguistics*, *14*(3), 220-229. https://doi.org/10.17509/ijal.v14i3.67603

## INTRODUCTION

Corpus linguistics (CL) is the study of language based on examples of "real-life" language use (McEnery & Wilson, 2001; Misnawati, et al., 2024) – large collections of authentic language data (Poole, 2018) – by looking at the language variations that can occur depending on the context in which it is used (Crawford & Csomay, 2016). CL informs vocabulary-based activities and dictionary production, enabling lexicographers to define words fully and explore their various meanings. It also informs vocabulary instruction through academic word lists used regularly in many English for Academic Purposes (EAP) settings (Friginal, 2018).

*Corresponding author
Email: misnawati_amir@yahoo.com

In short, CL provides insights into how people use language today across different situations.

While corpus linguistics has offered interesting possibilities for materials production (Anthony et al., 2017; Boulton & Tyne, 2015; Cheng, 2013; Dang, 2022; Du et al., 2021; Fraser et al., 2015) and is recognized as a potentially significant learning activity for students (Hou, 2014; Önen & İnal, 2019; Poole, 2022; Thurston & Candlin, 1998), this potential has not been fully realized in the classroom. Few specific corpus-based tools are explicitly created for didactic use, with the exception of general platforms like Kahoot or Web Sketchpad. Furthermore, a systematic literature review conducted by the researchers reveals that some areas in corpus-based studies remain underexplored, such as public administration, marketing, sociology, anthropology, and philosophy. Most researchers tend to focus on corpus linguistics applications in medicine (Fraser et al., 2015; Hsu, 2013; Lei & Liu, 2016), engineering (Jin, 2015; Jin et al., 2013; Ward, 2007), business (Hsu, 2011a, 2011b; Laosrirattanachai & Ruangjaroon, 2021; Yin & Li, 2021), and the social sciences (Chanasattru & Tangkiengsirisin, 2016; Mozaffari & Moini, 2014).

This gap is particularly significant in public administration, where students often grapple with the specialized vocabulary and complex language structures inherent in the field. A recent study by Misnawati et al. (2024) found that many students struggle to understand unfamiliar terminology in public administration, highlighting the need for targeted vocabulary acquisition. This linguistic barrier can hinder their comprehension of key concepts, policy analysis, and effective communication in professional settings. While numerous corpora exist, their application in public administration needs to be improved, leaving a significant gap in resources tailored to the specific needs of public administration learners (Meyer, 2023).

Existing studies by Nation (2004), Gardner (2007), and Coxhead (2000) have significantly contributed to corpus-based vocabulary research. However, their work needs to directly address the specific language demands of specialized domains like public administration. For instance, Coxhead's (2000) work utilizes the "word family" concept, which may not fully capture the specific terminology and phraseology crucial for navigating public administration discourse. Furthermore, while valuable for general language analysis, corpus development has chiefly employed quantitative analysis, such as corpus comparison (Nguyen Le & Miller, 2020; Quero, 2017; Zhang, 2013) or mixed methodologies (Browne, 2014). Research opportunities utilizing qualitative approaches (Nation, 2016), such as semantic scales, technical dictionaries, expert consultation, surveys, interviews, questionnaires, or classroom-based approaches, are relatively rare. Similarly, Brezina's (2012) purely quantitative approach may not adequately address the contextual nuances of specialized vocabulary in public administration. While insightful, these approaches often overlook the crucial role of qualitative analysis in ensuring the pedagogical relevance and user-friendliness of specialized corpora.

This study addresses these limitations by developing the English Corpus of Public Administration (ECOPA), a novel corpus-based didactic tool specifically designed to address the global linguistic challenges that public administration students and practitioners face. Unlike existing resources, ECOPA employs a rigorous qualitative methodology, incorporating semantic scales and expert consultation to ensure the corpus's relevance and pedagogical value. This research further contributes to the field by developing a public administration word list based on the word-lemma concept, prioritizing individual word forms over broader families to enhance precision in understanding public administration terminology. By providing a comprehensive and accessible repository of authentic language use in the field, ECOPA aims to enhance learners' vocabulary acquisition, improve their understanding of critical concepts, and ultimately contribute to more effective communication and practice in public administration.

**Corpus (*Plural: Corpora*)**

A corpus is a large, principled collection of naturally occurring language texts stored electronically (Reppen, 2010; Sinclair, 1991), encompassing both written and spoken forms (O'Keeffe & McCarthy, 2010). *"Naturally occurring texts"* refers to language derived from authentic language situations, such as conversations, meetings, letters, class assignments, and books, rather than fabricated or artificial language (Reppen, 2010). The collection must be *"principled"* to ensure it accurately represents the specific aspects of a language under investigation (Sinclair, 1991). In essence, a corpus is a vast collection of authentic language data, potentially gathered from sources like newspapers, blogs, academic essays, journals, and books, that has been compiled, organized, and made searchable. This allows for the creation of corpora focused on specific genres, such as student essays, political speeches, academic lectures, newspaper articles, and blogs (Poole, 2018). As Cheng (2012) emphasizes, a corpus is a collection of texts compiled for a particular reason, with representativeness being a key consideration.

While corpora are often associated with corpus linguistics, their use extends beyond this field. Although manually compiled corpora have existed for a long time, the advent of machine-readable

corpora has revolutionized the field. A good corpus is representative of the type of language it aims to capture (Biber, 1993). For example, if a corpus is designed to represent written language, the corpus designer must first identify the various written language situations (e.g., fiction, academic prose, personal letters) and then develop a strategy for collecting representative texts from each category (Biber et al., 2010; Friginal, 2018). A corpus can be general, containing a wide range of texts to provide a broad overview of a language, or specialized, focusing on a particular field like law, medicine, or public administration. In the latter case, the designer must collect texts (written or spoken) specifically related to that field. Corpus design involves careful consideration of factors such as representativeness, size, and balance to ensure the data effectively addresses the research questions (Boulton & Tyne, 2015). Ultimately, corpora contribute significantly to the identification and understanding of specialized vocabulary.

**Wordlist or Vocabulary in Corpus**

Wordlists are fundamental to effective vocabulary course design, the development of graded materials for extensive reading and listening, research on vocabulary load, and vocabulary test development (Nation, 2016). Traditionally, a wordlist is a vocabulary list organized by frequency or alphabetical order (McCarthy & O'Keeffe, 2010; O'Keeffe & McCarthy, 2010). More recently, wordlists have been constructed to identify terminology frequently used in specific genres, proving valuable in various linguistic studies, including language teaching (D. Gardner, 2007; Rungrueang et al., 2022).

Nation (2001) categorized vocabulary for English language learners into four types: (1) high-frequency words, essential general service words used frequently in everyday life and appearing across a wide range of contexts (Hsu, 2011a); (2) academic vocabulary, "sub-technical vocabulary" used across different academic disciplines (Coxhead & Nation, 2001); (3) technical vocabulary, specialized words used within a specific field, forming a technical wordlist for that discipline (Coxhead & Nation, 2001; Csomay & Petrovioc, 2012); and (4) low-frequency words, words with specific meanings within a particular field of expertise, occurring frequently in specialized texts but rarely elsewhere (Nation, 2001).

Vocabulary size, the number of words a person knows to some degree, is directly related to language learning proficiency across all skills (Miralpeix & Muñoz, 2018); it refers to the number of words learners know, at least at a basic level of meaning recognition (Wero et al., 2021). Estimates suggest that non-native English speakers need to understand at least 6,000-7,000-word families for spoken English and 8,000-9,000 for written English (Nation, 2006; Schmitt et al., 2017). Other research indicates that non-native speakers need to know at least 3,000-word families for basic conversation (Schmitt & Schmitt, 2014) and the most frequent 5,000-word families for movie comprehension (Webb & Rodgers, 2009).

**The Qualitatively Corpus Approach Development**

No single corpus is universally suitable for all purposes. The best corpus is the one that best meets the specific needs of the research at hand (O'Keeffe et al., 2007). Therefore, when compiling a wordlist from a corpus, several indicators should be considered (Nation, 2016): 1) Corpus type; 2) Corpus selection, including text types, geographical divisions, age-related material, and language learning situation; 3) Corpus size (word number); 4) Corpus proportion; 5) Corpus software/program; and 6) Corpus analysis.

In this study, the compiled corpus is a specialized corpus of written texts using US English, designed for teenagers or adults, and sourced from TV shows/movies and academic texts. It contains over 5,000 high-frequency words covering academic, narrative, and general writing. The corpus was compiled using AntConc and Antword Converter software, employing a qualitative corpus analysis approach with semantic scales and expert consultation.

To create a specialized corpus, Toriida (2016) proposes a multi-step process: 1) Investigating the target material, considering the teaching context, student needs, and intended corpus use; 2) Collecting the corpus based on the needs analysis; 3) Eliminating non-content words from the corpus; 4) Analyzing the corpus using text analysis software or manual analysis with expert assistance and specialized dictionaries; 5) Developing a frequency-based vocabulary listincluding part of speech, definitions, and sample sentences; 6) Validating the list with experts; and 7) Making final adjustments based on expert feedback.

Therefore, the research question guiding this study is: "How is the English corpus of public administration (ECOPA) developed through qualitative corpus analysis?"

## METHOD

**Research Design**

This study employs a qualitative approach within a corpus linguistics framework. Corpus linguistics involves the compilation and analysis of corpora (Cheng, 2013), encompassing a set of procedures for studying language based on real-life language use, utilizing both quantitative and qualitative methods (McEnery & Hardie, 2012; McEnery & Wilson, 2001). Qualitative corpus analysis, a specific methodology within this framework, enables in-

depth investigations of linguistic phenomena grounded in the context of language use. It involves systematically studying and analyzing language patterns within large collections of texts to gain qualitative insights and understand the nuances of language use (Hasko, 2012).

This study utilizes qualitative approaches, specifically expert consultation and semantic scales, for two primary reasons: 1) Expert Insight: Experienced lecturers in public administration provide valuable insights into learners' specific needs and vocabulary challenges, drawn from their direct interaction with students and the field. 2) Comprehensive Coverage: Expert input ensures the corpus encompasses a wide range of vocabulary and concepts relevant to learners, which is crucial for developing comprehensive word lists and effective language learning materials.

**Wordlist Creation**

Following Toriida's (2016) step-by-step guide for specialized corpus creation, this study involved: 1) Preparing the corpus: This included targeting relevant materials and eliminating unnecessary texts. 2) Analyzing the text using software (AntConc): This involved preparing the text for analysis in AntConc and generating a frequency list. 3) Developing an annotated frequency-based vocabulary list: This included identifying parts of speech, definitions, collocations, and sample sentences.

The initial step involved preparing the corpus by collecting materials related to 27 specific public administration topics (see Table 3). These topics were selected based on the curricula of universities with public administration programs, such as Cahaya Prima University, Yappi Makassar College of Administration, and Gorontalo University. Texts from 561 sources (available at https://www.ycit.or.id/ecopa/references.php) were collected from various online sources, including websites, journals, reference books, and textbooks, ensuring that the data was electronically stored and accessible to all corpus users. All references are documented on the ECOPA website, including titles, authors, URLs, and access dates.

Next, Antword Converter and AntConc (Anthony, 2022) were used to analyze the collected texts. All texts were converted into ".txt" files and processed in AntConc to identify word frequency, with a total of 24,226,582-word tokens. Following Nation's (2016) recommendation, words with a frequency above 50 were grouped and considered potential public administration terms. Before expert validation, a pre-validation stage was conducted to refine the wordlist. During this phase, researchers eliminated unnecessary or less relevant words, such as stop words, numbers, symbols, and typos.

The subsequent stage involved expert validation to finalize the English Corpus of Public Administration (ECOPA). Two expert validators, carefully selected for their expertise in public administration vocabulary, reviewed the 9,217 words (see Table 4) using Dang's (2020) semantic scale (see Table 1). Dang's (2020) framework was crucial in guiding the experts to identify general high-frequency words with specialized meanings in public administration, reflecting the vocabulary students are likely to encounter in their discipline (Coxhead & Demecheleer, 2018). The first validator was an English lecturer in a public administration program since 2014, and the second was a lecturer in a master's program in public administration since 2018, with a strong publication record in international journals. Their profound understanding of public administration terminology qualified them exceptionally for this validation task.

Finally, the ECOPA was compiled and uploaded to the project website. The researchers annotated the wordlist with parts of speech, definitions, and sample sentences, creating a specialized public administration wordlist. (More details on corpus construction are presented in the "Findings" section.) This ECOPA wordlist is a valuable resource for public administration learners, providing exposure to authentic language use and facilitating vocabulary acquisition. Lecturers can use ECOPA to design language-focused activities and assessments, while scholars can leverage it for further research into public administration language.

**Table 1**

*Semantic scale used for experts' validation (Dang, 2020, p.12)*

| Scale | Description |
|---|---|
| 1 | The word has no relationship with public administration. |
| 2 | The word has a meaning related to public administration and is (almost) the same as the meaning in everyday language use. |
| 3 | The word has a meaning related to public administration and differs from everyday language use. |
| 4 | The word has only one (or more) meaning(s), and it is (they are) only related to public administration. |

During the validation process, when experts were uncertain about assigning a semantic scale rating to a word, concordance lines were provided to aid their decision-making. Words receiving a rating of one from both experts were removed from the list, while those rated 2, 3, or 4 were retained as public administration terms. Here is a step-by-step guide to facilitate this process:

**Table 2**
*The corpus analysis procedures*

| Steps | Descriptions |
|---|---|
| Collecting Target Material | *Select Relevant Texts:* Choose texts and materials relevant to public administration, such as textbooks, journal articles, novels, graded readers, course materials, and movie scripts that cover the necessary topics.<br>*Gather Materials:* Collect the texts. This involved sourcing physical copies, downloading digital versions, or obtaining materials from libraries or online resources.<br>*Digital Conversion:* Convert all materials into a digital format suitable for analysis. Scan physical documents into PDF or Word documents, ensuring all texts are readable and accessible for processing. |
| Corpus Analysis | *Review Content:* Carefully review the collected materials to identify and remove any irrelevant or unnecessary text for the corpus study, focusing only on content that directly contributes to understanding the vocabulary needs.<br>*Cleanse Data:* To streamline the analysis process, ensure the corpus is free from unneeded texts. This may involve deleting text sections, removing duplicates, or excluding off-topic materials. |
| Experts Validation | *Identify Experts:* Choose two lecturers or experts with extensive knowledge of public administration vocabulary.<br>*Provide Materials:* Share the ECOPA, concordance lists, and the semantic scale with the experts. Ensure they have access to all necessary tools and documents for a thorough review.<br>*Explain Criteria:* Clarify the criteria on the semantic scale, explaining how it guides the inclusion or exclusion of words and the evaluation of their relevance to public administration studies.<br>*Concordance List Review:* Instruct the experts to use the concordance lists to examine the context and usage of words, aiding in their decisions about what should be included or excluded.<br>*Withdraw the Data:* After the initial review, withdraw the data of the experts' findings and any discrepancies in their evaluations.<br>*Wordlist Decision:* Decide which words to include or exclude based on the semantic scale results. The goal is to ensure the corpus accurately reflects vocabulary essential for EPA students.<br>*Proofreading:* Have the experts proofread the final selection of words to ensure accuracy, proper context, and relevance. |
| Finalization of ECOPA | *Final Adjustments:* Based on the experts' feedback, make any necessary adjustments to the corpus.<br>*ECOPA launch:* Launch the ECOPA by making it available to the target audience, such as public administration educators and students. |

## FINDINGS

The following findings detail the practical application of the corpus development methodology outlined in the previous section. Each step, from data collection and processing to validation and finalization, is presented with its outcomes, demonstrating a clear alignment between the planned procedures and the actual development of ECOPA.

### Collecting Target Material

The "collecting target material" process was pivotal in constructing ECOPA. This process required a detailed and systematic strategy for material aggregation, essential for the corpus's subsequent development and utility. The initial phase involved an exhaustive examination of data relevant to the public administration field. The researcher embarked on an extensive gathering of specific resources, with the selection of topics guided by the curricula of universities with public administration programs, such as Cahaya Prima University, Yappi Makassar College of Administration, Gorontalo University, and Hasanuddin University.

**Table 3**
*Topics collected for ECOPA development*

| ECOPA Topics | | |
|---|---|---|
| Public Administration | Principles of Management | Leadership |
| Development Administration | Strategic of Management | Family Sociology |
| Tax Administration | Human Resources Management | Political Science |
| Business Administration | Basic Sciences of Organization and Management | Introduction to Socio-cultural Anthropology |
| Administrative Analysis | Demography and Development | Introduction to Sociology |
| Administrative Ethics | Regional Development Administration Issues | Introduction to Social Statistics |
| Administrative Behavior | Policy Issues | Economic System |
| Comparison of State Administration | Public policy | Information and Management Systems |
| Political Science | Public Service Management | Government and Society |

As seen in Table 3, the 27 themes or topics were carefully selected and categorized to encompass a broad spectrum within public administration and its related fields. This included fundamental areas like public administration, development administration, and tax administration, as well as specialized areas like administrative analysis and administrative ethics. The collection also included materials on administrative behavior, state administration comparison, political science, various management disciplines, demography, development, regional administration issues, policy matters, public policy, public service management, and leadership studies. Furthermore, the corpus was expanded to include literature on economic, information, and management systems, and the interactions between government and society.

The researchers emphasized gathering diverse texts, including storytelling, general knowledge, and academic literature (formal and informal). A total of 561 resources were utilized, sourced from websites (66.48%), books (23.71%), textbooks/modules (3.39%), and journal articles (6.42%). The texts were compiled by searching for the intended topics through Google, which led to resources on Wikipedia, Google Scholar, Google Books, university websites, and online course sites.

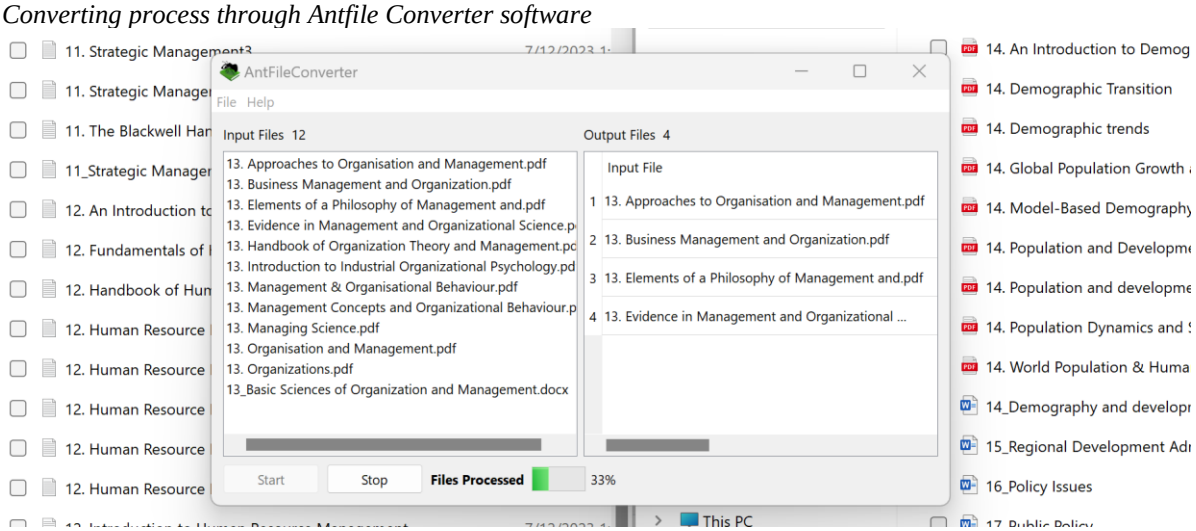
The next crucial stage involved carefully processing and examining the collected materials. This stage went beyond simple data collection; it required a detailed analysis to determine the corpus's relevance and usefulness to public administration. The aim was to create a corpus that was comprehensive and tailored to the language needs of public administration students, transforming ECOPA into a valuable instructional resource.

**Corpus Analysis**

Corpus analysis was as crucial as collecting the target material in developing ECOPA. This section details the procedures undertaken to convert the 561 sources on public administration into a unified and analyzable collection. The initial approach entailed thorough profiling and proofreading of the sources to ensure data accuracy and relevance, a crucial step in constructing a reliable database for analysis. Text from websites was meticulously transcribed and organized into Microsoft Word documents, while scholarly articles and other resources in PDF format were downloaded. This initial phase was crucial in establishing a well-organized and readily available database.

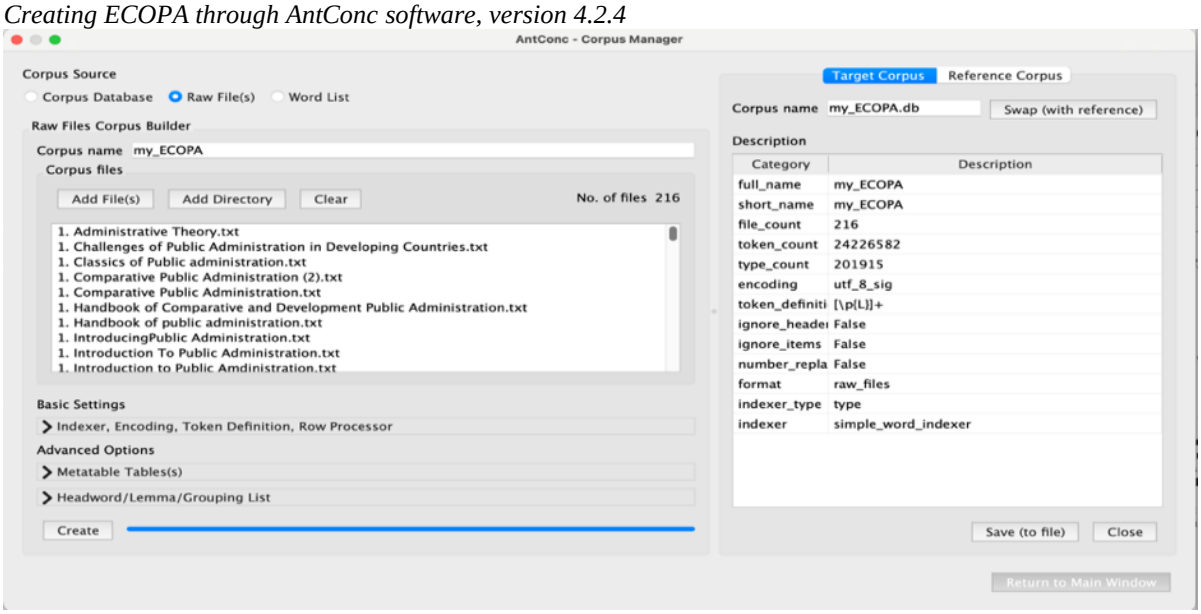
**Figure 1**

*Converting process through Antfile Converter software*



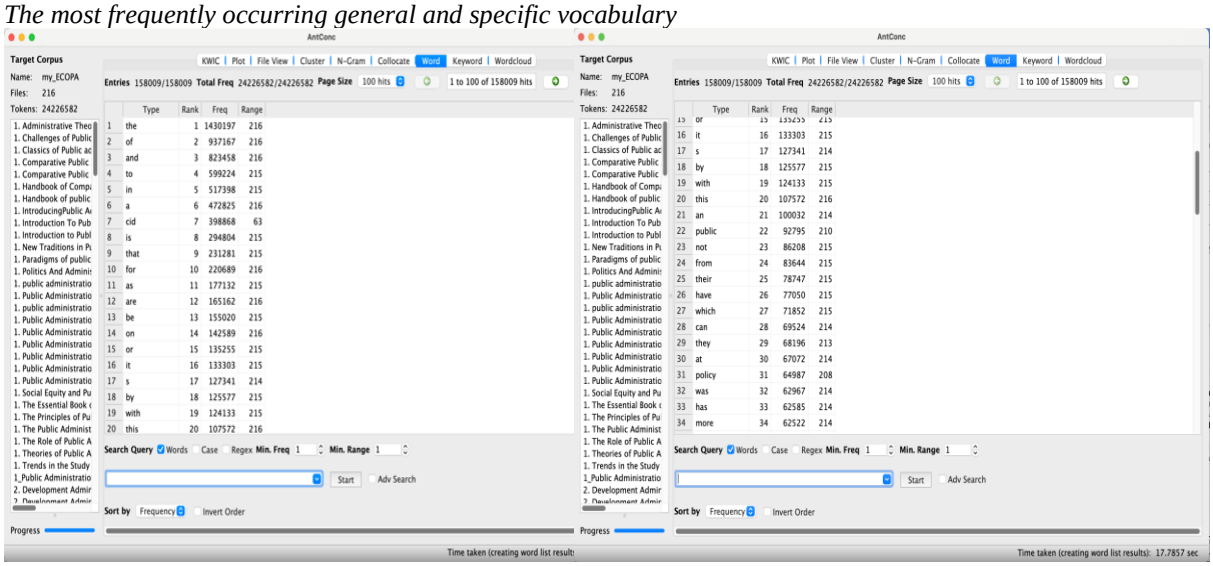

In the next phase, all materials were converted into ".txt" format using Antfile Converter to ensure compatibility with AntConc, the chosen software for corpus analysis (Figure 1). This conversion involved merging the 561 resources into 216 documents (Word and PDF files). This step simplified the analysis process and ensured consistency in the corpus format, reducing potential inconsistencies in data processing.

**Figure 2**
*Creating ECOPA through AntConc software, version 4.2.4*




After converting all files to ".txt" format, the researcher used AntConc software (version 4.2.4). The 216 text files, collectively named "my ECOPA," were uploaded to the software for efficient handling and searching (see Figure 2). This step uncovered a large dataset consisting of 24,226,582 million tokens, providing a basis for understanding the corpus's vast extent.

**Figure 3**
*The most frequently occurring general and specific vocabulary*




The initial frequency analysis (Figure 3) revealed that in the 24,226,582 million tokens, the word "the" was the most frequent general vocabulary word (1,430,197 occurrences), while "public" was the most frequent term specifically related to public administration (92,795 occurrences). This frequency analysis informed the creation of ECOPA's wordlists. A list of 158,009 words was generated from AntConc, with 141,473 words occurring less than 50 times (<50) and 16,536 words occurring 50 times or more (≥50). Words with <50 occurrences were eliminated from the pre-validation wordlist, while words with ≥50 occurrences were retained.

**Figure 4**
*Process of developing the ECOPA*


After the frequency study, the researcher began creating ECOPA by considering both frequency and concordance (Figure 4). The pre-validation process systematically removed general vocabulary items, resulting in a narrowed list of 9,217 words (see Table 4). This ensured the corpus focused on the most relevant and commonly used terms in public administration.

Before validation, the corpus was annotated, including grammatical tagging (classifying each word by its part of speech) and semantic tagging (providing detailed word meanings). The annotations helped validators verify the suitability and linguistic precision of each term, ensuring the corpus was academically and practically useful for individuals involved in public administration.

Finally, ECOPA underwent a thorough validation process by public administration experts to ensure the contextual relevance and accuracy of each term. This validation process was essential in guaranteeing the corpus's precision and dependability, increasing its value as a resource.

**Expert Validation**

Following corpus analysis and design, expert validation was crucial for verifying the accuracy and relevance of ECOPA's content. ECOPA employed a qualitative corpus analysis method (Coxhead, 2000) involving a semantic scale technique. This method, while time-consuming, ensured comprehensive corpus validation.

The validation process utilized a four-level semantic scale (Dang, 2020), ranging from 1 (no connection to public administration) to 4 (exclusive use in public administration contexts). Terms rated 1 by both experts were eliminated from the corpus, while terms rated 2, 3, or 4 were retained. This process ensured the corpus included only relevant terminology. When validators faced challenges classifying a word, a concordance list was created to provide additional context and aid decision-making.

**Table 4**
*Total number of words in the pre-validation stage*

| Scale | Number of words |
|---|---|
| 1 (The word has no relationship with public administration) | 2.933 |
| 2 (The word has a meaning related to public administration and is (almost) the same as the meaning in everyday language use) | 4.445 |
| 3 (The word has a meaning related to public administration and is different from the meaning in everyday language use) | 1.673 |
| 4 (The word has only one (or more) meaning(s), and it is (they are) only related to public administration) | 166 |
| Word counts | 9.217 |

Table 4 displays the word count assigned to each scale after the researchers' pre-validation. These 9,217 words were compiled into a validation sheet and submitted to two expert validators. The validators were selected for their extensive expertise in public administration vocabulary. One was an English lecturer in a public administration program since 2014, and the other was a lecturer in a master's program in public administration since 2018, with a strong publication record.

The validation sheet provided to the validators initially categorized 2.933 words as "Scale 1" during pre-validation. This was followed by 4.445 words categorized as "Scale 2," 1.673 words categorized as "Scale 3," and 166 words categorized as "Scale 4." The initial categorization provided a foundation for the validators, who had the authority to either agree with this initial classification or assign words to different scales based on their expert assessment. The validators included words escalated in scale in the ECOPA list, whereas words downscaled to scale 1 were excluded.

**Table 5**
*ECOPA validation process*

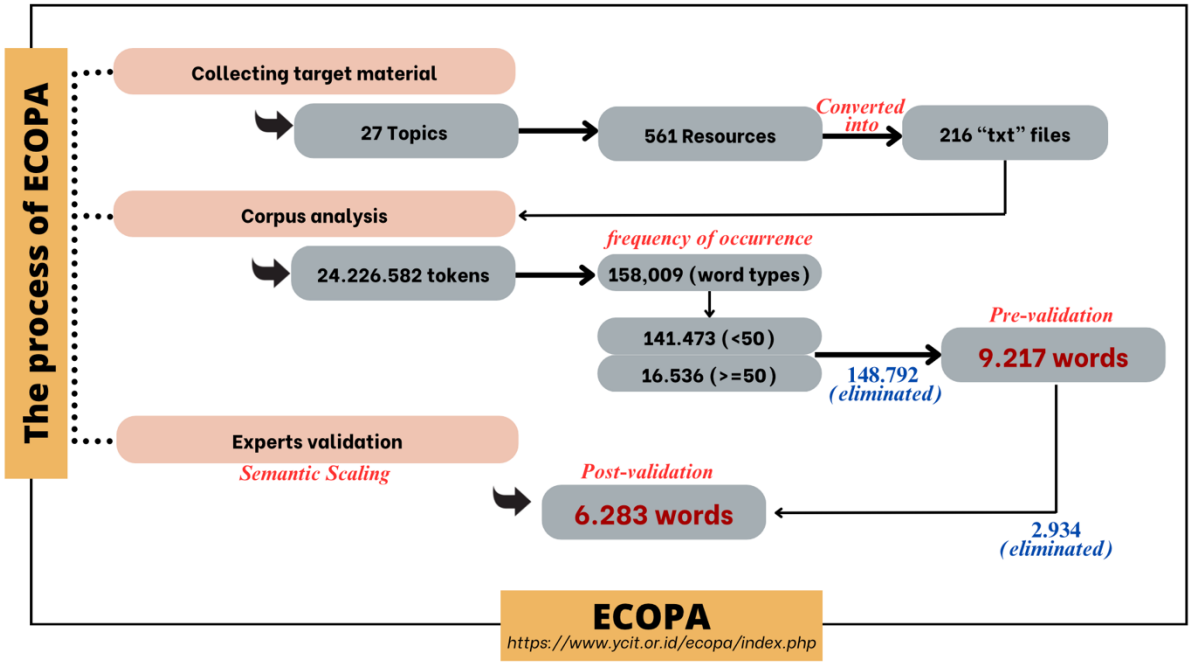
| Scale | Pre - Validation | Validator 1 | Validator 2 | Final Validation | ECOPA |
|---|---|---|---|---|---|
| 1 | 2.933 | 2.936 | 2.934 | 2.934 | - |
| 2 | 4.445 | 4.444 | 4.444 | 4.444 | 4.444 |
| 3 | 1.673 | 1.673 | 1.673 | 1.673 | 1.673 |
| 4 | 166 | 164 | 166 | 166 | 166 |
| Word counts | 9.217 | 9.217 | 9.217 | 9.217 | 6.283 |

Following the validation process (Table 5), ECOPA was finalized with 6,283 words representing the specific vocabulary of public administration. This curated compilation ensures ECOPA's reliability and value for those studying and working in the field. The validation process was crucial in developing a comprehensive corpus aligned with the discipline's current language needs.

The validation process also included a feedback system where validators could offer insights and comments, which were carefully reviewed and incorporated into the final corpus when appropriate. This iterative feedback and refining process significantly improved ECOPA's quality and accuracy.

As general findings, Figure 5 provides a flow diagram illustrating the word counts throughout the ECOPA development process, from the initial collection to the final validated list, clearly visualizing the steps and outcomes.

**Figure 5**
*A flow diagram of the ECOPA process*



**English Corpus of Public Administration (ECOPA)**

The creation of ECOPA represents significant progress in English language acquisition and instruction in public administration. ECOPA, a product of thorough research and analysis, serves as a didactic tool for educators and students in public administration programs.

ECOPA has both indirect and direct applications. Indirectly, it is a valuable resource for developing dictionaries and educational materials. Its inclusion in teaching English for public administration enhances the educational experience by offering readily available and contextually relevant linguistic tools. ECOPA helps instructors create well-informed and contextually relevant instructional materials, improving the overall standard of teaching.

Directly, ECOPA serves as a primary data source for linguistic analysis, allowing students and instructors to interact actively with the corpus, generating insights and learning opportunities. This engagement promotes a deeper understanding of language use in public administration and enhances experiential learning by allowing users to apply their language skills in authentic contexts.

**Figure 6**

*ECOPA website*


To optimize accessibility, ECOPA is available online at https://www.ycit.or.id/ecopa/index.php. Figure 6 shows the website's user-friendly interface, making it easy for users to navigate and access the corpus.

Figure 7 shows the comprehensive list of references used to compile ECOPA, allowing users to access the original materials directly. This provides transparency and showcases the diverse range of texts included in ECOPA, from general and academic to formal and informal, making it a versatile resource.

**Figure 7**

*References page on the ECOPA website*

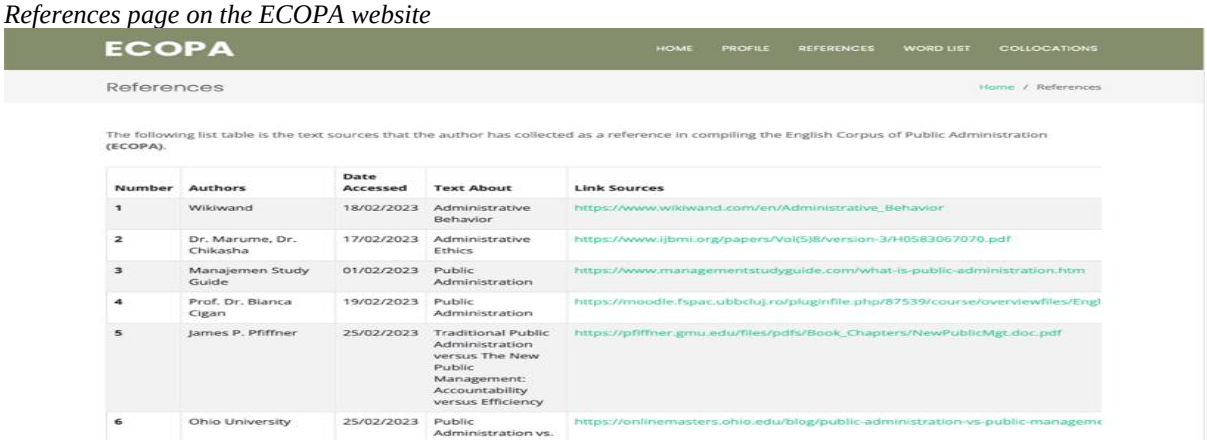
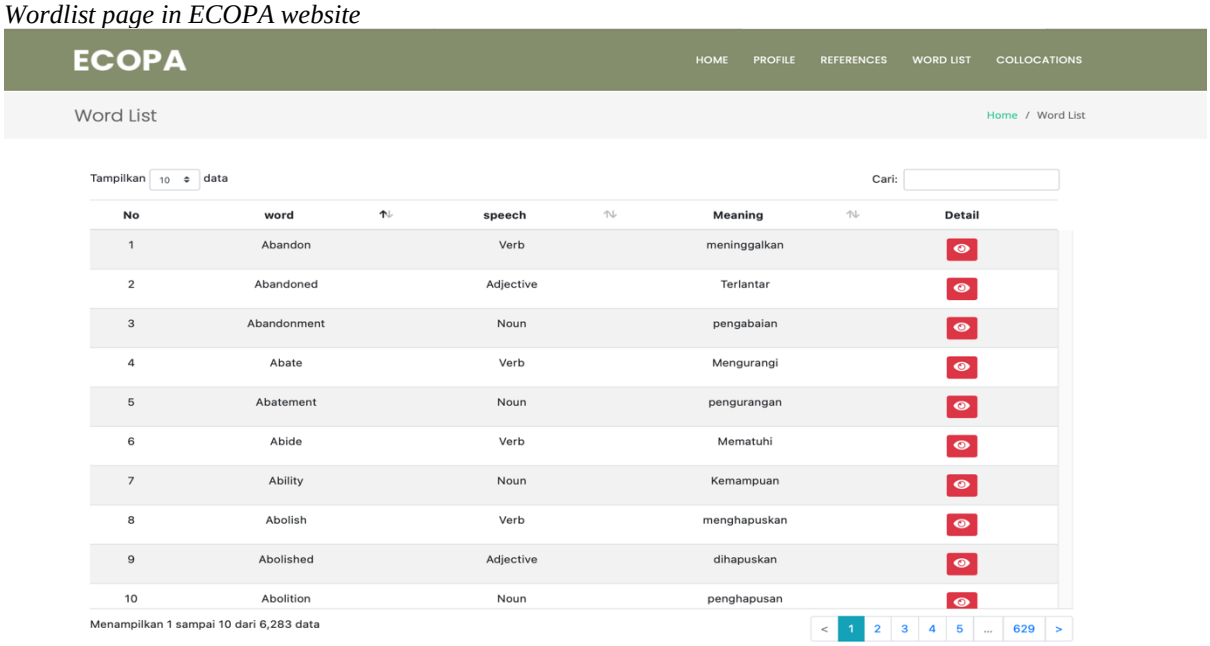


**Figure 8**
*Wordlist page in ECOPA website*




The "wordlist" page is a crucial element, showcasing the validated collection of 6,283 terms relevant to public administration. This page provides information on each word, including its part of speech, meaning, and concordance. The concordance feature offers real-life examples of how each word is used in the field, enhancing the learning experience. Figure 8 displays a list of terms beginning with the letter 'A,' including the concordance for the word "public."

In summary, the creation and implementation of ECOPA represent significant progress in English language instruction in public administration. ECOPA, a dynamic and interactive resource based on empirical evidence, enhances the learning experience for students and enriches teaching practices for educators. It serves as a comprehensive and readily available collection of language data, contributing to the evolving field of language education.

## DISCUSSION

Developing ECOPA demonstrates a meticulous and methodical approach to researching public administration terminology. The development process involved four key steps: target material collection, corpus analysis, corpus validation, and ECOPA launch.

First, collecting target material for ECOPA is crucial for improving vocabulary and language skills among public administration students. The method used to construct ECOPA involved systematic strategies for gathering materials, essential for the corpus's development and usefulness. This adheres to the principles of corpus linguistics, which prioritizes the significance of natural language usage in developing linguistic resources (Stefanowitsch, 2020). Lee et al. (2019) and McEnery (2019) have examined the importance of corpora in understanding language use, which is essential for developing curricula that address learners' specific vocabulary needs. Đurović (2021) advocated for using corpora to produce dictionaries and teaching resources, highlighting the importance of authentic language examples in educational settings.

The researchers began by comprehensively gathering specialized resources encompassing various themes and topics relevant to public administration. This collection covered fundamental and specialized areas within the discipline, aligning with the literature on domain-specific language acquisition. Gardner (2021), Murray and Christison (2019) investigated the importance of understanding and acquiring language used in specific academic disciplines, emphasizing the need for students and professionals to interact effectively within their fields. This supports the comprehensive strategy used to gather a wide range of resources for ECOPA, ensuring the collection represents the broad scope of language use in public administration (Egbert et al., 2020). The diverse sources provided a variety of texts, including narrative, general, academic, formal, and informal literature, encompassing websites, books, textbooks, and journal articles.

The next crucial step involved carefully processing and examining the gathered materials during the corpus analysis phase. This went beyond mere data collection and included a detailed content analysis to determine the corpus's relevance and suitability for public administration. The objective was to create a comprehensive and refined

collection of texts that met the linguistic needs of public administration students. This aligns with recommended methods for designing and analyzing corpora (Friginal & Hardy, 2020; Fuster-Márquez et al., 2020; Reppen & Simpson-Vlach, 2019). These scholars emphasize the importance of ensuring the corpus is representative and balanced to meet the target audience's needs.

Second, conducting corpus analysis is crucial for understanding the precise linguistic requirements of ECOPA. Meticulously profiling and proofreading sources are essential for developing a structured and reliable database for analysis. Antfile Converter software was used to convert materials into ".txt" format, ensuring compatibility with AntConc software for analysis. This ensured consistency in the corpus structure. Paquot and Gries (2021), Adamou (2019), Meyer (2023), and Đurović (2021) highlight the importance of transforming texts into a standardized format for analysis. Their work examines techniques in corpus linguistics, including frequency and concordance analysis, to understand linguistic patterns.

The corpus analysis uncovered a large dataset with notable differences in the frequency of general and specific terminology, providing insights into the linguistic characteristics within the corpus. The development of ECOPA involved analyzing frequency and concordance data, leading to a precise compilation of relevant terms in public administration. AntConc software (Anthony, 2022) facilitated these comprehensive analyses, providing practical insights and enabling the evaluation of frequency data and the identification of key terms (Alamri, 2022).

The annotation process, which included grammatical and semantic labeling, was essential for ensuring the clarity and usability of the corpus. It enhanced comprehension of the language and vocabulary relevant to public administration. The final validation process, conducted by public administration experts, confirmed the corpus's accuracy and trustworthiness, increasing its value as a resource for students, educators, and practitioners. This process aligns with the work of Barth and Schnell (2021), Dash (2021), Newman and Cox (2021), and Rayson and Chapelle (2019), who highlight the importance of grammatical and semantic tagging to improve a corpus's clarity and usefulness. This comprehensive step, involving thorough frequency analysis, meticulous annotation, and validation, generated a linguistically sound collection of data for practical purposes, enhancing the understanding and utilization of language in public administration.

Third, validating the corpus was a crucial aspect of ECOPA's development. After corpus analysis, this step ensured the accuracy and relevance of the content. Biber and Reppen (2015) emphasize the critical need for expert validation in ensuring a corpus's accuracy and dependability. For ECOPA, this process involved engaging specialists to authenticate the content, increasing its value for educational and practical purposes. The process employed a qualitative corpus analysis approach, incorporating a semantic scale method for comprehensive validation.

The semantic scale used for validation ranged from levels denoting no connection to exclusive utilization in public administration contexts. Less relevant words were consistently eliminated, while those more relevant to public administration terminology were kept. This refined the corpus to include only relevant terms. Tognini-Bonelli (2001) presented the concept of qualitative corpus analysis, highlighting the importance of context and semantic scales in understanding word meaning. This methodology is consistent with the ECOPA validation process, which used a semantic scale to ascertain the significance of words in relation to public administration.

The validation procedure evaluated each word's semantic and contextual significance, considering its frequency and relevance in public administration contexts. This involved lexical analysis and scrutinizing usage circumstances to ensure an accurate portrayal of public administration terms, aligning with Baker's (2006) research on using corpora for discourse analysis. This methodology was essential in evaluating the semantic and contextual significance of terms in ECOPA, ensuring the corpus accurately reflected the vocabulary used in public administration.

Finally, ECOPA serves as a primary data source for language analysis, enabling students and educators to interact directly with the corpus and extract significant insights. This interaction promotes a deeper and more practical understanding of language use in public administration. This aligns with the work of Johns (1991), Ma and Mei (2021), and Szudarski (2023), who advocated for utilizing concordance data from corpora to enhance language acquisition. This approach allows learners to explore language patterns and usage, consistent with the direct application of ECOPA for linguistic analysis and learning. The creation and implementation of ECOPA represent a significant advancement in language teaching, providing a dynamic and engaging learning experience for students in public administration based on empirical evidence. It also enhances teaching practices for educators in this field. ECOPA is a comprehensive and accessible corpus that reflects the evolving landscape of language education and serves as a vital tool in public administration.

## CONCLUSION

This study makes a significant contribution to the field of public administration by developing the

English Corpus of Public Administration (ECOPA), a novel corpus-based didactic tool. Unlike existing resources, ECOPA addresses the need for a comprehensive and accessible repository of authentic language use in public administration. The rigorous methodology employed—combining Toriida's (2016) corpus analysis steps with Dang's (2020) semantic scale—ensures the corpus's reliability and validity for both academic and practical applications. The resulting compilation of 6,283 online terms serves as an invaluable resource for learners, educators, and researchers seeking to deepen their understanding of public administration language.

While this study successfully achieved its goal of developing a specialized corpus, certain limitations highlight opportunities for future research. The study's reliance on written texts may not fully capture the nuances of spoken language in public administration contexts. Future research could explore incorporating spoken corpora to bridge this gap. Additionally, while expert validation ensured content relevance, further investigation into specific pedagogical applications of ECOPA—such as the development of targeted language learning materials or tools for assessing language proficiency—would enhance its practical utility. Lastly, the study acknowledges the assumption of digital literacy among its users. Future research could examine the impact of varying levels of digital literacy on ECOPA's accessibility and effectiveness, potentially leading to the creation of supplementary resources or tailored user guides.

**ACKNOWLEDGEMENTS**

The first author extends her sincere appreciation to the Indonesian Education Scholarship (BPI) under the auspices of the Indonesia Endowment Fund for Education Agency (LPDP) and the Education Service Centre (Puslapdik), a division of the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek). Their support has been instrumental in facilitating her doctoral studies from 2021 to 2024.

**REFERENCES**

Adamou, E. (2019). Corpus linguistic methods. In *Handbücher zur Sprach- und Kommunikationswissenschaft* [Handbooks of Linguistics and Communication Science] (pp. 638–653). De Gruyter. https://doi.org/10.1515/9783110435351-052

Alamri, B. (2022). The role of corpus linguistics in grammar instruction: A review of literature. *International Journal of Linguistics*, *14*(6), 158–167.

Anthony, L. (2022). *AntConc (Version 4.1.4)*. Waseda University. https://www.laurenceanthony.net/software

Anthony, L., Chen, M., & Flowerdew, J. (2017, July 23). Introducing in-service English language teachers to corpus-assisted academic writing pedagogy: A Hong Kong case. *Corpus Linguistics 2017 International Conference*. https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper259.pdf

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Barth, D., & Schnell, S. (2021). *Understanding corpus linguistics*. Routledge.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257.

Biber, D., & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics* (Vol. 40). Cambridge University Press. https://doi.org/10.1515/icame-2016-0009

Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed, pp. 548–570). Oxford University Press.

Brezina, V. (2012). Use of Google Scholar in corpus-driven EAP research. *Journal of English for Academic Purposes*, *11*(4), 319–331. https://doi.org/10.1016/j.jeap.2012.08.001

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction*, *3*(1), 1–10. https://doi.org/10.7820/vli.v03.1.browne

Chanasattru, S., & Tangkiengsirisin, S. (2016). Developing of a high-frequency word list in Social Sciences. *Journal of English Studies*, *11*, 41–87. https://so04.tci-thaijo.org/index.php/jsel/article/view/73309

Cheng, W. (2012). *Exploring corpus linguistics: Language in action* (First). Routledge.

Cheng, W. (2013). Corpus-based linguistic approaches to critical discourse analysis. *The Encyclopedia of Applied Linguistics*, 1353–1360.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/https://doi.org/10.2307/3587951

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes*, *51*, 84–97. https://doi.org/10.1016/j.esp.2018.03.006

Coxhead, A., & Nation, P. (2001). *The specialized vocabulary of English for academic purposes*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9781139524766.020

Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics* (First publ). Routledge.

Csomay, E., & Petrovioc, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System*, *40*, 305–315. https://doi.org/10.1016/j.system.2012.05.004

Dang, T. N. Y. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline related TV programs: Insights from medical corpora. In *TESOL Quarterly* (Vol. 54, Issue 2, pp. 436–459). https://doi.org/10.1002/tesq.552

Dang, T. N. Y. (2022). A corpus-based study of vocabulary in conference presentations. *Journal of English for Academic Purposes*, *59*, 101144. https://doi.org/https://doi.org/10.1016/j.jeap.2022.101144

Dash, N. S. (2021). *Language corpora annotation and processing*. Springer.

Du, Z., Jiang, F., & Liu, L. (2021). Profiling figure legends in scientific research articles: A corpus-driven approach. *Journal of English for Academic Purposes*, *54*, 101054. https://doi.org/https://doi.org/10.1016/j.jeap.2021.101054

Đurović, Z. (2021). Corpus linguistics methods for building ESP word lists, glossaries, and dictionaries on the example of a Marine Engineering word list. *Lexikos*, *series 31*, 259–282. https://doi.org/https://doi.org/10.5788/31-1-1647

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Fraser, S., Davies, W., & Tatsukawa, K. (2015). Creating a corpus-informed EMP course for medical undergraduates. *Journal of the IATEFL ESP SIG*, *45*.

Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities* (1st ed). Routledge.

Friginal, E., & Hardy, J. A. (2020). *The Routledge handbook of corpus approaches to discourse analysis*. Routledge.

Fuster-Márquez, M., Gregori-Signes, C., & Ruiz, J. S. (2020). *Multiperspectives in analysis and corpus design*. Editorial Comares.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*(2), 241–265. https://doi.org/https://doi.org/10.1093/applin/amm010

Gardner, H. (2021). *Disciplined mind: What all students should understand*. Simon & Schuster.

Hasko, V. (2012). Qualitative corpus analysis. In *The Encyclopedia of Applied Linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0974

Hou, H.-I. (2014). Teaching specialized vocabulary by integrating a corpus-based approach: Implications for ESP course design at the university level. *English Language Teaching*, *7*(5), 26–37.

Hsu, W. (2011a). A business word list for prospective EFL business postgraduates. *The Asian ESP Journal*, *7*(4), 63-79.

Hsu, W. (2011b). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes*, *30*(4), 247–257. https://doi.org/10.1016/j.esp.2011.04.005

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454-484. https://doi.org/10.1177/1362168813494121

Jin, N. Y. (2015). *Development of a corpus of Malaysian KBSM engineering texts and related word list*. University Putra Malaysia.

Jin, N. Y., Ling, L. Y., Tong, C. S., Sahiddan, N., Philip, A., Azmi, N. H. N., & Tarmizi, M. A. A. (2013). Development of the engineering technology word list for vocational schools in Malaysia. *International Education Research*, *1*(1), 43–59. https://doi.org/10.12735/ier.v1i1p43

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. *English Language Research Journal*, *4*, 1–16.

Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(1), 50–86. https://ir.swu.ac.th/jspui/handle/123456789/17519

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753. https://api.semanticscholar.org/CorpusID:149388749

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. https://doi.org/10.1016/j.jeap.2016.01.008

Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, *1*(1), 177–190. DOI: 10.1515/jccall-2021-2008

McCarthy, M., & O'Keeffe, A. (2010). Historical perspective what are corpora and how have they evolved? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3–13). Routledge.

McEnery, T. (2019). *Corpus linguistics*. Edinburgh University Press.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (Second Edi). Edinburgh University Press ltd.

Meyer, B. (2023). Corpus-based studies of public service interpreting 1. In *The Routledge Handbook of Public Service Interpreting* (pp. 76–88). Routledge.

Meyer, C. F. (2023). *English corpus linguistics: An introduction*. Cambridge University Press.

Miralpeix, I., & Muñoz, C. (2018). *Receptive vocabulary size and its relationship to EFL language skills*. *56*(1), 1–24. https://doi.org/doi:10.1515/iral-2017-0016

Misnawati, M., Atmowardoyo, H., Sulaiman, I., Yusriadi, Y., & Rahman, A. (2024). Unveiling the lecturers' and students' needs in English for public administration program: Essential vocabulary topics, instructional methods, and learning challenges. *Register Journal*, *17*(1), 100–122. https://doi.org/10.18326/rgt.v17i1.100-122

Misnawati, M., Nur, S., & Tahir, S. Z. (2024). Corpus linguistics today: A qualitative approach. *Research and Innovation in Applied Linguistics [RIAL]*, *2*(1), 45–62. https://doi.org/10.31963/rial.v2i1.4486

Mozaffari, A., & Moini, R. (2014). Academic words in education research articles: A corpus study. *Procedia - Social and Behavioral Sciences*, *98*, 1290–1296. https://doi.org/10.1016/j.sbspro.2014.03.545

Murray, D. E., & Christison, M. (2019). *What English language teachers need to know volume I: Understanding learning*. Routledge.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. *Vocabulary in a Second Language*, 3–13.

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

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company. https://doi.org/10.1075/z.208

Newman, J., & Cox, C. (2021). Corpus annotation. In *A practical handbook of corpus linguistics* (pp. 25–48). Springer.

Nguyen Le, C. N., & Miller, J. (2020). A corpus-based list of commonly used English medical morphemes for students learning English for specific purposes. *English for Specific Purposes*, *58*, 102–121. https://doi.org/https://doi.org/10.1016/j.esp.2020.01.004

O'Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics* (First Edition). Routledge. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). *From corpus to classroom: language use and language teaching*. Cambridge University Press.

Önen, S., & İnal, D. (2019). A corpus-driven analysis of explicitness in English as lingua franca. *Journal of Curriculum and Teaching*, *8*(3), 73–83. https://doi.org/10.5430/jct.v8n3p73

Poole, R. (2018). *A guide to using corpora for English language learners*. Edinburgh University Press Ltd.

Poole, R. (2022). "Corpus can be tricky": revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, *35*(7), 1620–1641. https://doi.org/10.1080/09588221.2020.1825095

Quero, B. (2017). A corpus comparison approach for estimating the vocabulary load of medical textbooks using the GSL, AWL, and EAP science lists. *TESOL International Journal*, *12*(1), 177–192. https://eric.ed.gov/?id=EJ1247827

Rayson, P., & Chapelle, C. (2019). Corpus analysis of key words. *The Concise Encyclopedia of Applied Linguistics*, 320–326. https://doi.org/10.1002/9781405198431.wbeal0247.pub2

Reppen, R. (2010). *Using corpora in the language classroom* (1st ed.). Cambridge University Press.

Reppen, R., & Simpson-Vlach, R. (2019). Corpus linguistics. In *An introduction to applied linguistics* (pp. 91–108). Routledge.

Rungrueang, T., Boonprasert, P., Boonprasert, P., & Boonprasert, P. (2022). Corpus-based approach to generate a word list for food service. *THAITESOL Journal*, *35*(1), 57–76. https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/258591

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb

(2007). *Language Teaching*, *50*(2), 212–226. https://doi.org/DOI: 10.1017/S0261444815000075

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

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

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

Szudarski, P. (2023). *Collocations, corpora and language learning*. Cambridge University Press.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*(3), 267–280. https://doi.org/10.1016/S0889-4906(97)00013-6

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.6

Toriida, M.-C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list. *TESL Canada Journal/Revue TESL Du Canada*, *34*(11), 87–105. https://doi.org/http://dx.doi.org/1018806/tesl.v34i1.1255

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, *6*(1), 18–35. https://doi.org/10.1016/j.jeap.2006.10.001

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Wero, Y. T., Machmud, K., & Husain, N. (2021). The study on students' vocabulary size. *Jambura Journal of English Teaching and Literature*, *2*(1), 22–34. https://doi.org/10.37905/jetl.v2i1.10279

Yin, X., & Li, S. (2021). Lexical bundles as an intradisciplinary and interdisciplinary mark: A corpus-based study of research articles from business, biology, and applied linguistics. *Applied Corpus Linguistics*, *1*(1), 100006. https://doi.org/https://doi.org/10.1016/j.acorp.2021.100006

Zhang, M. (2013). A corpus-based comparative study of semi-technical and technical vocabulary. *The Asian ESP Journal*, *9*(2 Spesial Edition), 148–172. http://asian-esp-journal.com/wp-content/uploads/2016/01/AESP-Volume-9-Issue-2-October-2013.pdf