



**PAPER – OPEN ACCESS**

## Exploring the Use of the Term 'Pemilu' in the Context of the 2024 General Election in Indonesia: A Corpus Linguistics Study

Author : T. Thyrhaya Zein et al  
DOI : 10.32734/lwsa.v7i2.2095  
Electronic ISSN : 2654-7066  
Print ISSN : 2654-7058

*Volume 7 Issue 2 – 2022 TALENTA Conference Series: Local Wisdom, Social, and Arts (LWSA)*



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

Published under licence by TALENTA Publisher, Universitas Sumatera Utara



# Exploring the Use of the Term “Pemilu” in the Context of the 2024 General Election in Indonesia: A Corpus Linguistics Study

T. Thyrhaya Zein\*, Dwi Kurnia Surya Ningrum, T. Silvana Sinar, Christanta Rejuna Phanes S. Brahmana

Universitas Sumatera Utara, Medan, Indonesia

t.thyrhaya@usu.ac.id

## Abstract

This research employs a corpus linguistics approach to conduct a comprehensive analysis of language usage in the context of the Indonesia's 2024 General Election (Pemilu). Corpus linguistics is an approach that utilizes language corpora as data sources to investigate patterns and meanings embedded in the term "pemilu" and its collocations. The study aims to uncover linguistic nuances relevant to significant events like the 2024 General Election and their alignment with the United Nation's Sustainable Development Goals (SDGs). The research method involves corpus linguistics analysis, which systematically examines language data to identify usage patterns within the context of the 2024 General Election. The study collects Indonesian language corpus data from 100 texts that taken from diverse online news sites, preprocesses the data, and employs the CQPweb software for frequency and collocation analysis. The research findings show that the term "pemilu" appears 887 times in the corpus used, with a frequency of 17,885.949 occurrences per million words. The findings also highlight collocations such as "2024", "tahap" (stages), "peserta" (participants), "kampanye" (campaign), "hasil" (results), "DPT" (Final Voter List), "2019", "sistem" (system), "penyelenggara" (organizer), and "UU" (legislation), shedding light on key aspects and themes related to the 2024 General Election. The interconnections between these collocates offer valuable insights into Indonesia's electoral discussions and their alignment with the SDGs, particularly SDG 16's focus on inclusive, accountable, and just societies.

**Keywords:** collocations; corpus; general election; SDGs

## Abstrak

Penelitian ini menggunakan pendekatan linguistik korpus untuk melakukan analisis komprehensif terhadap penggunaan bahasa dalam konteks Pemilihan Umum (Pemilu) Indonesia tahun 2024. Linguistik korpus merupakan suatu pendekatan yang memanfaatkan korpora bahasa sebagai sumber data untuk menyelidiki pola dan makna yang terkandung dalam istilah "pemilu" dan kolokasinya. Studi ini bertujuan untuk mengungkap nuansa linguistik yang relevan dengan peristiwa penting seperti Pemilu 2024 dan keselarasan mereka dengan Tujuan Pembangunan Berkelanjutan (SDGs) Perserikatan Bangsa-Bangsa. Metode penelitiannya melibatkan analisis linguistik korpus, yang secara sistematis mengkaji data bahasa untuk mengidentifikasi pola penggunaan dalam konteks Pemilu 2024. Penelitian ini mengumpulkan data korpus berbahasa Indonesia dari 100 teks yang diambil dari berbagai situs berita online, memproses data terlebih dahulu, dan menggunakan perangkat lunak CQPweb untuk analisis frekuensi dan kolokasi. Hasil penelitian menunjukkan bahwa istilah "pemilu" muncul sebanyak 887 kali dalam korpus yang digunakan, dengan frekuensi 17.885.949 kemunculan per juta kata. Temuan ini juga menyoroti kolokasi seperti "2024", "tahap", "peserta", "kampanye", "hasil", "DPT" (Daftar Pemilih Tetap), "2019", "sistem", "penyelenggara", dan "UU" (undang-undang), yang menyoroti aspek dan tema utama terkait Pemilu 2024. Keterhubungan antara kolokasi ini memberikan wawasan berharga mengenai diskusi pemilu di Indonesia dan keselarasan mereka dengan SDGs, khususnya fokus SDG 16 pada masyarakat yang inklusif, akuntabel, dan adil.

**Kata Kunci:** kolokasi; korpus; pemilihan umum; SDG

## 1. Introduction

General Elections (Pemilihan Umum or Pemilu) in Indonesia are crucial political events held periodically, providing a platform for political parties and aspiring leaders to win the hearts of voters and attain power. Language plays a significant role in influencing and connecting with voters during this electoral process. However, the use of language in the political context has become increasingly complex, particularly in the digital era. Texts in various forms, such as online news, social media posts, or survey reports, have proliferated and diversified. Therefore, a study of language in the political context has become increasingly important, especially in the lead-up to the next General Election, where it aligns with the United Nations' Sustainable Development Goals (SDGs).

The SDGs serve as a global framework to address pressing challenges, foster inclusive, accountable, and participatory governance (SDG 16), and achieve a sustainable future for all. In Indonesia, the General Elections (Pemilu) play a pivotal role in determining the leadership of the nation and its legislative bodies, influencing policies that impact society's well-being and the nation's progress towards SDG-aligned governance.

Corpus linguistics is a field of study that employs a set of methods for analyzing language using machine-readable texts, known as corpora (McEnery & Hardie, 2011). This approach utilizes data from language materials collected in a source called a corpus, which includes texts from various genres, varieties, and both oral and written forms. The corpus data is systematically organized and easily accessible electronically through computers. Tognini-Bonelli (2001) delineates two corpus linguistics approaches: corpus-based and corpus-driven. They differ in their view of the corpus as evidence that can support a theory.

The corpus-based approach is deductive, while the corpus-driven approach is inductive, considering the corpus as evidence that should guide the theory. Corpus-based research possesses four main characteristics: (1) Empirical analysis based on actual usage patterns; the corpus comprises various communication activities, such as collections of texts from books, fiction, non-fiction, magazines, papers, literature, newspapers, phone conversations, mobile phone conversations, meetings, lectures, radio broadcasts, and television programs. (2) Utilization of large and structured text collections as the basis for analysis; this characteristic refers to the corpus itself, which can take the written corpus, spoken corpus, and more. (3) Integration of technology in the analysis; technology aids not only in storing corpus but also in assisting the analysis of language within the corpus. The corpus can be accessed and analysed using concordance programs available on computers. (4) Combination of quantitative and qualitative analysis; the quantitative results from the corpus will be qualitatively analysed to draw research conclusion.

This paper centers on corpus analysis, based on Baker (2010) claim that corpus linguistics entails examining typically extensive sets of electronically stored texts with the assistance of computer software. Computers have the capability to swiftly and accurately calculate frequencies and conduct statistical tests, providing researchers with insights into linguistic patterns and trends. Corpus analysis empowers researchers to validate or challenge hypothesis regarding language use, and it also enables them to pose new inquiries and theories about language that would be otherwise unfeasible. As Crawford & Csomay (2019) mentioned that a corpus functions as a representative compilation of language, offering insights into language use. In the context of the 2024 General Election in Indonesia, corpus linguistics can serve as an effective tool for analysing language usage and its impact on sustainable development. It involves using a corpus of the Indonesian language as a data source, which comprises a collection of language texts gathered and stored digitally for linguistic analysis. Through corpus linguistics, researchers can systematically and objectively analyse language usage, identifying patterns within the context of the 2024 General Election, while aligning with the principles of inclusive governance, accountability, and civic engagement outlined in SDG 16.

In this study, collocations of the term “pemilu” within the context of the 2024 General Election in Indonesia will be sought. Stubbs (2001) defines collocation as the co-occurrence of words. It involves a node in the form of a word-form that is the lemma or focus being investigated, accompanied by collocates that co-occur with the lemma in the corpus. In collocation studies, the main focus is to discover patterns of association between frequently appear together with the lemma, known as collocates, can be identified. By studying collocation patterns, we can understand the relationship between words used in specific contexts. This analysis provides insights into linguistic structures, word meanings, and the contextual usage of words in the corpus texts under investigation.

However, the use of corpus linguistics to analyse language usage in the context of the 2024 General Election is still relatively rare, particularly concerning its connection to the SDGs. Therefore, this research aims to analyse the Indonesian language corpus within the context of the 2024 General Election using the corpus linguistics approach and explore its implications for sustainable development. The researcher will employ CQP Web, software for analysing language corpora that has proven effective in conducting corpus analyses.

CQPweb was chosen as the corpus analysis tool for this study due to its versatility, efficiency, and user-friendly interface. Unlike some other corpus analysis tools, CQPweb offers a web-based platform that allows researchers to access and analyse language corpora without need for complex installation or extensive technical expertise. Its user-friendly interface enables efficient querying and exploration of linguistic patterns, making it accessible for researchers with varying levels of experience in corpus linguistics. CQPweb also provides powerful analytical features, such as concordance analysis and collocation identification, which are crucial for extracting meaningful insights from large datasets. Additionally, CQPweb’s ability to handle both written and spoken language corpora ensures compatibility with a diverse range of textual sources. By utilizing CQP Web, the researcher will analyse the Indonesian language corpus from various online news sites.

Several previous studies have discussed the use of language in various contexts such as political, medical, educational, and others, using corpus analysis from diverse media sources. For instance, Artha (2018) revealed the representation of the Indonesian government (from the era of the New Order to the post-reformation period) in the COCA (Corpus of Contemporary American English) and COHA (Corpus of Historical American English). The researcher uncovered how the image of the Indonesian government is portrayed through adjectives associated with the word “Indonesian.” By employing a corpus-driven Critical Discourse Analysis approach based on corpus linguistics, the researcher identified distinct patterns of adjective usage for each government period, and these variations in adjective usage represented the image of the Indonesian government.

In a study by Purwitarini (2020), the semantic prosody of the words “suspect” and “patient” in COVID-19 corpora was analysed. The research aimed to identify their frequency of usage and significant collocates, indicating positive or negative tendencies. The study applied the qualitative method with the COVID-19 data instrument through a sketch engine and used a corpus approach. The findings stated that “Patient” was found to be more widely used in COVID-19 scientific research, with positive collocations, while “suspect” had negative collocations. This study highlights how language can influence perceptions in health crises.

The study conducted by Yuliawati et al. (2018) explored the terminology of the social sciences and humanities in Indonesian scientific papers using a corpus-based approach. They integrated corpus linguistic methods with the communicative theory of terminology (CTT) to examine terminological units. Through keyword and word clusters analysis, they identified multi-word terms, while collocation analysis revealed significant sense categories of terms. The study demonstrated how the corpus-based approach offers benefits in compiling, describing, and presenting terms more reliably and efficiently. It also provides an alternative method for creating glossaries and resolving terminological problems for translators (Rosa et al., 2020).

The research conducted by Oktavianti & Ardianti (2019) explored a corpus-based analysis of verbs in the news section of The Jakarta Post, focusing on the frequency of verbs and their relationship with the characteristics of news text. The study compiled a corpus of news articles from October to December 2018, comprising a total of 21,682 words. The researchers categorized verb types based on Scheibmann’s compilation of Halliday’s verb taxonomy and Dixon’s verb types. The analysis revealed that verbal verbs were the most frequent, followed by material and existential verbs, while corporeal and perception/relational verbs were the least frequent. The research suggested a strong correlation between the frequency of verbs in news text and the text’s characteristics. Verbs that aligned with the function of delivering information were more prevalent, while those not aligned with the text’s purpose were infrequently used.

Another research by Ramayanti et al. (2022) delved into a corpus-based analysis of the adjectives “adorable,” “charming,” and “cute” in Corpus of Contemporary American English (COCA). The study aimed to uncover and analyse the collocative meanings of these three adjectives, which are often used interchangeably due to their shared similar meanings. The data were collected from COCA, employing a documentation method with note-taking techniques. The study utilized the theory of collocative meaning and the theory of lexical collocation for its corpus-based analysis. The findings demonstrated that the collocative meanings of the three adjectives were frequently associated with nouns like “baby,” “guy,” “boys,” and others. Additionally, they were commonly paired with adverbs such as “quite,” “so,” “always,” and more.

Laliy Zen (2020) utilized a corpus-driven approach to analyze language patterns in children's literature using the CoCL (Corpus of Children's Literature), comprising 28 novels and short stories from Project Gutenberg. The study focused on works by Carlo Collodi, Lewis Carroll, Beatrix Potter, and Hugh Lofting, comparing the CoCL's 319,968 tokens with the BNC Written Imaginative. Findings highlighted distinctive features in children's literature, such as increased use of nouns and subjective pronouns, explicit themes of smallness, animals, and food, and the cultivation of positive and joyful tones, reflecting optimism. The paper contributes to understanding the role of corpora in linguistic and literary analysis, showcasing the impact of the digital world on language research.

However, there is a research gap when it comes to understanding how language usage influences political discourse and communication within the specific context of the 2024 General Election in Indonesia. While previous studies have explored language in various contexts like governmental representation, health crises, semantic prosody, scientific papers, and news article analysis, there hasn't been a direct investigation into the language patterns used in an upcoming general election and their implications for sustainable development.

This research intends to bridge this gap by employing the corpus linguistics method to analyse language patterns within the context of the 2024 General Election in Indonesia. The primary focus will be on depicting crucial aspects of the election process, such as election year dynamics, stages, the roles of election organizers, and insight drawn from past election experiences. These aspects collectively contribute to comprehensive understanding of the 2024 General Election within the studied corpus.

The significance of this research lies in its focus on language usage within upcoming 2024 General Election in Indonesia, it apart from previous studies that explored diverse contexts such as governmental representation, scientific papers and news articles. Unlike retrospective analysis, this study provides real-time insights into language dynamics surrounding a pivotal event. This research is expected to provide in-depth and comprehensive insights into language usage in the context of the 2024 General Election, while shedding light on its connections to the principles of inclusive and accountable governance outlined in SDG 16. Furthermore, the results of this study can contribute to the development of corpus linguistics in Indonesia, offering valuable perspectives to enhance political discourse and communication, and promote a more informed and engaged electorate towards achieving the SDGs.

The selection of this topic is informed by several compelling factors. Firstly, general elections hold immense political significance, shaping the nation’s leadership and future trajectory. Investigating the language employed in these elections offers insight into the communication strategies of political parties, candidates, and voters. Secondly, the digital age has revolutionized political discourse, with online platforms and social media playing a pivotal role in shaping public opinion. Thirdly, by aligning the research with the Sustainable Development Goals (SDGs), the study takes on a broader societal context. It examines how political language intersects with the principles of inclusivity, accountability, and sustainable development outlined in the SDGs. Moreover, this research contributes to relatively underexplored area of linguistic analysis, focusing on political language in Indonesia’s specific context. By employing the rigorous methodology of corpus linguistics, the study offers an empirical understanding of language usage in political discourse, potentially guiding effective communication strategies. Ultimately, this investigation holds the potential to enhance democratic processes by deciphering the role of language in shaping perceptions, fostering public engagement, and advancing sustainable development ideals.

**2. Method**

The research method utilized in this study is corpus linguistics analysis, as defined by Sudaryanto (2015) as a systematic approach to examining language data. This method is employed to analyse Indonesian language corpus data from various online news sites with the aim of identifying language usage patterns within the context of the 2024 General Election.

The initial stage of this research method involves collecting Indonesian language corpus data. The corpus data will be sourced from various online news sites, such as detik.com, kompas.com, CNN Indonesia, and Antara News. There are 100 news texts collected from May – August 2023. The collected corpus data will be in digital format and will serve as raw data for subsequent corpus analysis.

The second stage encompasses the pre-processing of the corpus data. This phase involves selecting relevant data, as well as grouping words and processing tokens. The objective of this stage is to facilitate the subsequent corpus data analysis.

The third stage encompasses the pre-processing of the corpus data. This phase involves corpus data analysis using the CQPweb Lancaster University corpus analysis website. CQPweb, which stands for Corpus Query Program, is a robust web-based tool for querying corpora, enabling users to apply various techniques in Corpus Linguistics (CL) analysis. This web-based application allows for cross-platform usage, eliminating the need for users to install or index corpora on their local PCs even though such functionality is also present in CQPweb Lancaster. This platform is employed to conduct frequency analysis of words, collocation analysis, and concordance analysis on the corpus data. Frequency analysis identifies frequently used words in the context of the 2024 General Election.

Collocation analysis identifies the words that commonly appear together with specific words, while concordance analysis examines the contextual usage of specific words within the corpus data. This stage also involves the removal or elimination of noise, such as unwanted words before move to the next stage.

The final stage consists of interpreting of the results of the corpus data analysis and formulating conclusions. Upon obtaining the results of the corpus analysis, the researchers will interpret these findings and draw conclusions from the study. These conclusions will provide answers to the research questions and objectives. Throughout the research process, ethical considerations will be taken into account, including the use of valid data and copyright adherence. Additionally, the validity of data will be ensured throughout data collection, analysis, and interpretation phases.

**3. Results**

The query “pemilu” conducted in the CQP Web yielded 887 matches across 91 distinct texts (within 49,592 words [100 texts]; frequency: 17,885.949 instances per million words). The outcomes are visualized in the captured concordance and frequency breakdown screenshots, providing a comprehensive view of the contextual occurrences of the term “pemilu.” The frequency details for the term “pemilu” are illustrated below.

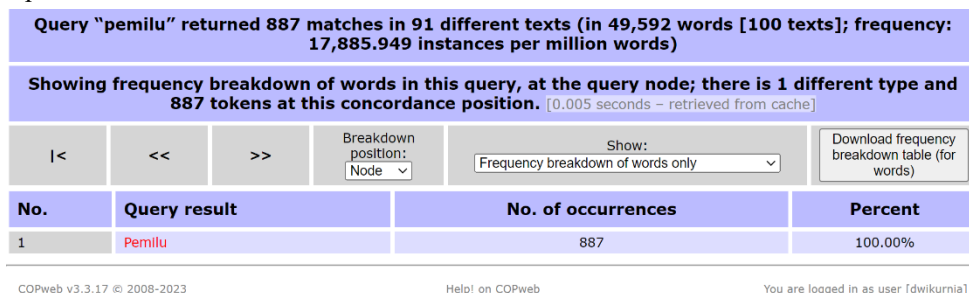


Fig. 1. Frequency details of the term “pemilu”

From the information about concordance provided above, it is understood that the term “*pemilu*” appeared 887 times in 91 different texts within the corpus. The total words in the analysed corpus amount to 49,592 words derived from 100 texts. The frequency of the term “*pemilu*” is calculated at 17,885.949 instances per million words. This frequency metric demonstrates how often the term “*pemilu*” occurs within every million words of the corpus, shedding light on the term’s prevalence and prominence in the corpus data.

After obtaining the concordance data for the term “*pemilu*” from the corpus, the next step is to identify the collocations of the word “*pemilu*,” which are presented in the table and figure below:

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log Ratio (filtered)
1	pemilu	7	0.751	7	2	6.864
2	peserta	85	9.122	79	25	5.775
3	mempertanyakan	5	0.537	5	3	6.378
4	lanjutan	5	0.537	5	2	6.378
5	monev/pengawasan	5	0.537	5	4	6.378
6	larangan	5	0.537	5	2	6.378
7	pantauan	33	3.541	29	9	5.914
8	UU	54	5.795	47	11	5.804
9	eksekutif	7	0.751	6	3	5.641
10	mewanda	7	0.751	6	3	5.641
11	selaku	5	0.644	5	2	5.378
12	penyelenggara	57	6.117	47	12	5.289
13	2019	66	7.083	54	18	5.226
14	jelang	10	1.073	8	5	5.095
15	keutusan	10	1.073	8	3	5.095
16	rena	9	0.966	7	6	4.864
17	merajang	24	2.576	18	11	4.941

Fig. 2. Collocations for the term “*pemilu*”

The collocation span is 3LR. This means CQPweb will consider 3 words to the left and right sides of “*pemilu*”, as the collocate candidates. The filter set by the minimum observed collocate of 5 words. This means words which occur less than 5 times within the collocation will not be considered. This filter is used to exclude words with low frequency. The information provided in the figure above indicates that there are 1,722 distinct word types present in the collocation database for the query “*pemilu*.” Out of those 1,722 different word types, only 42 distinct words have been sorted and presented in the table below as collocations with the term “*pemilu*.”

Table 1. Collocations of the term “*pemilu*”

No.	Word	Total no. in whole corpus	Expected collocate frequency	Log Ratio (filtered) value	Observed collocate frequency	In no. of texts
1	2024	489	52.477	326	82	4.056
2	dalam	438	47.004	104	39	1.373
3	pada	284	30.478	73	39	1.525
4	tahapan	125	13.415	73	25	3.546
5	peserta	85	9.122	79	25	6.775
6	kampanye	84	9.015	34	8	2.500
7	hasil	74	7.941	23	14	1.907
8	DPT	72	7.727	24	6	2.056
9	2019	66	7.083	54	18	5.226
10	sistem	62	6.654	26	5	2.587
11	penyelenggara	57	6.117	47	12	5.289
12	UU	54	5.795	47	11	5.804
13	Badan	53	5.688	24	22	2.783
14	tentang	52	5.580	19	15	2.260
15	soal	44	4.722	15	7	2.105
16	pelaksanaan	39	4.185	27	12	4.226
17	Pilkada	39	4.185	27	9	4.226
18	serentak	38	4.078	27	9	4.352

19	Penyelenggaraan	38	4.078	25	7	4.000
20	mendatang	34	3.649	14	14	2.542
21	surat suara	34	3.649	15	3	2.715
22	penundaan	33	3.541	29	9	5.914
23	ikut	30	3.220	13	9	2.669
24	Pengawas	28	3.005	21	17	4.641
25	tantangan	27	2.898	12	3	2.734
26	menjelang	24	2.576	18	11	4.641
27	reporter	23	2.468	11	11	2.931
28	2017	20	2.146	12	12	3.641
29	proporsional	19	2.039	11	2	3.516
30	mengundurkan	17	1.824	10	2	3.571
31	D	11	1.181	7	6	3.864
32	jelang	10	1.073	8	5	5.056
33	kecurangan	10	1.073	8	1	5.056
34	news	9	0.966	7	6	4.864
35	menunda	7	0.751	6	3	5.641
36	putuskan	7	0.751	7	2	6.864
37	eksploitasi	7	0.751	6	1	5.641
38	selaku	6	0.644	5	2	5.378
39	menyelenggarakan	5	0.537	5	4	6.378
40	mengkampanyekan	5	0.537	5	3	6.378
41	lanjutkan	5	0.537	5	2	6.378
42	larangan	5	0.537	5	2	6.378

The data above consists of 42 collocation instances of the term “*pemilu*” that have been sorted by their frequency from the highest to the lowest. The subsequent step involves the process of selecting relevant data by eliminating “noise.” In the context of corpus linguistics analysis, as highlighted by Hunston & Francis (2000), “noise” refers to irrelevant data or unwanted words or data that do not contribute meaningfully to the analysis. This can encompass words lacking relevance or significant meaning to the topic under study. During the pre-processing phase of corpus data, the removal or elimination of noise is carried out to clean the data from irrelevant information, allowing the analysis to concentrate on more pertinent aspects aligned with the research objective. Through the elimination of noise, corpus data analysis can achieve heightened accuracy and significance, empowering researchers to identify clearer and more profound language usage patterns within a specific context. In table above, the data highlights in red are considered noise and will be removed from the data analysis.

Out of the initially collected 42 data sets that underwent pre-processing and noise reduction, a subset of 10 collocations data points has been chosen for detailed examination and discussion within this paper. These selected collocation instances represent the most frequently occurring combinations of words that co-occur with the term “*pemilu*.” These specific collocation instances are deemed particularly significant for analysis due to their high frequency of appearance within the corpus data. By focusing on these 10 prominent collocation examples, the paper aims to delve deeper into the linguistic relationships and patterns that emerge when the term “*pemilu*” is used in conjunction with other words. Here are the top 10 collocations of the term “*pemilu*,” displayed in the table below:

Table 2. Top 10 collocations of the term “*pemilu*”

No.	Word	Total no. in whole corpus	Expected collocate frequency	Log Ratio (filtered) value	Observed collocate frequency	In no. of texts
1	2024	489	52.477	326	82	4.056
2	Tahapan	125	13.415	73	25	3.546
3	Peserta	85	9.122	79	25	6.775
4	Kampanye	84	9.015	34	8	2.500
5	Hasil	74	7.941	23	14	1.907
6	DPT	72	7.727	24	6	2.056

7	2019	66	7.083	54	18	5.226
8	System	62	6.654	26	5	2.587
9	Penyelenggara	57	6.117	47	12	5.289
10	UU	54	5.795	47	11	5.804

Table 2 displays the top 10 collocations of the term “*pemilu*.” The data includes information on the total number of occurrences of each collocate in the entire corpus, the expected collocate frequency based on the corpus size, the observed collocate frequency, and the number of texts in which the collocate appears.

1. The word "2024" has a total of 489 occurrences in the entire corpus. Its expected collocate frequency is 52.477, but it has an observed collocate frequency of 326 in 82 texts. The Log Ratio (filtered) value of 4.056 indicates that "2024" is significantly associated with the term "*pemilu*," suggesting discussions and references related to the upcoming 2024 General Election.
2. The word "*tahapan*" (stages) appears 125 times in the corpus, with an expected collocate frequency of 13.415. Its observed collocate frequency is 73 in 25 texts, resulting in a Log Ratio (filtered) value of 3.546. This indicates that discussions about the stages of the election process are significantly linked to the term "*pemilu*."
3. The term "*peserta*" (participants) is observed 85 times in the corpus, exceeding the expected collocate frequency of 9.122. With an observed collocate frequency of 79 in 25 texts, it has a high Log Ratio (filtered) value of 6.775. This suggests a strong connection between the participants of the election and the term "*pemilu*."
4. "*Kampanye*" (Campaign) is observed 84 times in the corpus, with an expected collocate frequency of 9.015. It has an observed collocate frequency of 34 in 8 texts, resulting in a Log Ratio (filtered) value of 2.500. This indicates a moderate association between election campaigns and the term "*pemilu*."
5. The word "*hasil*" (Results) appears 74 times, with an expected collocate frequency of 7.941. Its observed collocate frequency is 23 in 14 texts, leading to a Log Ratio (filtered) value of 1.907. This suggests some connection between election results and the term "*pemilu*."
6. "*DPT*" (Final Voter List) is observed 72 times, exceeding the expected collocate frequency of 7.727. Its observed collocate frequency is 24 in 6 texts, resulting in a Log Ratio (filtered) value of 2.056. This indicates that discussions related to "*Daftar Pemilih Tetap*" (Permanent Voter List) are associated with the term "*pemilu*."
7. The term "2019" is observed 66 times, with a much higher expected collocate frequency of 7.083. It has an observed collocate frequency of 54 in 18 texts, resulting in a Log Ratio (filtered) value of 5.226. This suggests that discussions about the previous 2019 General Election are significantly linked to the term "*pemilu*."
8. "*Sistem*" (System) appears 62 times in the corpus, with an expected collocate frequency of 6.654. Its observed collocate frequency is 26 in 5 texts, leading to a Log Ratio (filtered) value of 2.587. This indicates some connection between election systems and the term "*pemilu*."
9. The term "*penyelenggara*" (Organizer) is observed 57 times, exceeding the expected collocate frequency of 6.117. With an observed collocate frequency of 47 in 12 texts, it has a high Log Ratio (filtered) value of 5.289. This suggests a significant connection between election organizers and the term "*pemilu*."
10. "*UU*" (Legislation) is observed 54 times, with a higher expected collocate frequency of 5.795. Its observed collocate frequency is 47 in 11 texts, resulting in a Log Ratio (filtered) value of 5.804. This indicates a strong association between legislation and the term "*pemilu*."

These collocations provide insights into the significant terms frequently associated with "*pemilu*" within the analysed corpus, shedding light on key aspects and themes related to the 2024 General Election.

#### 4. Discussion

The presented research findings reveal several intriguing collocates of the term “*pemilu*” (general election) worth discussing. Firstly, the word “2024” stands out as a high frequently occurring collocate alongside “*pemilu*” in the analysed corpus. The fact that this term appears 326 times in the analysed corpus, compared to an expected frequency of 52.477, suggests that the upcoming election year is a central focus of discussions within the collected texts. The presence of this collocation in 82 distinct texts reinforces its prominence across a variety of sources.

This finding holds implications for the broader socio-political landscape and its relation to the Sustainable Development Goals (SDGs). The upcoming 2024 General Election can play a pivotal role in contributing to SDGs, particularly SDG 16, which focuses on promoting peaceful, just, and inclusive societies. Free and fair elections, as a fundamental component of democratic governance, align with SDG 16's aims of accountable institutions, participatory decision-making, and access to justice. As such, the extensive discussions related to the 2024 General Election may signify a commitment to democratic values and inclusive governance that contribute to sustainable development.



Additionally, the prominence of discussions about the election year "2024" can reflect efforts to ensure stability and adherence to democratic processes, which are crucial for achieving various SDGs. The election's potential influence on policy-making, governance effectiveness, and public participation in decision-making processes can contribute to positive outcomes in areas such as poverty reduction, quality education, gender equality, and more.

The prominent collocation of "2024" with the term "*pemilu*" in this study, indicating its high frequency and widespread presence across diverse sources, contrast with Artha's (2018) examination of the Indonesian government's representation over historical periods. In this study, the frequent occurrence of "2024" suggests a concentrated focus on the upcoming election year, underscoring its significance in discussions. This is a departure from Artha's historical lens, highlighting the dynamic nature of language use in contemporary political contexts.

The frequent collocation of the word "*tahapan*" (stages) with the term "*pemilu*" (election) suggests a strong association between discussions about the stages of the election process and the concept of elections in the analysed corpus. The query "*tahapan*" is observed 125 times, with observed collocate frequency of 73 instances in 25 texts, and high Log Ratio (filtered) value of 3.546, indicates that the term "*tahapan*" frequently appears in contexts discussing various stages of the election process.

This finding holds potential implications in the context of the SDGs, particularly in the context of SDG 16. The term "*tahapan*," which refers to stages, indicates a structured process inherent in democratic elections. This structured approach aligns with the principles of SDG 16, as it underscores the significance of transparent and accountable electoral processes, citizen participation, and effective institutions.

The frequent occurrence of the collocation "*tahapan*" could signify that discussions about the different stages of the election process are a significant part of the analysed corpus. These discussions may encompass topics like candidate and voter registration, campaign periods, polling days, vote counting, and result announcement. Such discussions contribute to ensuring a fair and transparent electoral process, which is integral to SDG 16's goal of promoting peace, justice, and strong institutions. Therefore, the strong collocation of "*tahapan*" with "*pemilu*" highlights the significance of structured and accountable electoral processes in Indonesia's pursuit of sustainable development goals.

The collocation of "*peserta*" (participants) in discussions related to the election process indicates its significant relevance and use in describing individuals or entities involved in the electoral context. With an observed frequency of 79 times, surpassing the expected frequency of 9.122, and appearing in 25 distinct texts, it underscores the importance of highlighting the key factors contributing to the election process.

From the perspective of the Sustainable Development Goals (SDGs), this finding reflects a fundamental aspect of democratic governance and inclusive participation. SDG 16 emphasizes the promotion of inclusive and accountable institutions at all levels, and the participation of "*peserta*" aligns with this goal. In the context of elections, involving a wide range of participants ensures that diverse voices and perspectives are taken into account during the decision-making process. Inclusion of various stakeholders, including marginalized and underrepresented groups, contributes to a more comprehensive and equitable democratic process, thereby advancing SDG's 16 aim.

Furthermore, the emphasis on "*peserta*" can also relate to other SDGs. For instance, SDG 5 focuses on gender equality and the empowerment of women. Ensuring the active participation of women as "*peserta*" in the electoral process can contribute to bridging gender gaps and promoting women's representation in political affairs. Similarly, the inclusion of marginalized communities as active "*peserta*" (participants) can address issues of social inequality and promote social justice, in line with SDG 10.

The presence of the term "*kampanye*" (campaign) in the collocates of "*pemilu*" (election) indicates a moderate association between discussion about election campaigns and the overall concept of elections in the analysed corpus. The occurrence of the query "*kampanye*" for 84 times and the observed collocate frequency of 34 instances in 8 texts, along with a Log Ratio (filtered) value of 2.500, suggests that "*kampanye*" is moderately linked to the term "*pemilu*."

In the context of the SDGs, election campaigns hold significance for several reasons. SDG 16 (Peace, Justice, and Strong Institutions) emphasizes the importance of transparent, accountable, and inclusive institutions. Political campaigns are a critical component of the democratic process, allowing political candidates and parties to communicate their platforms and engage with citizens. Effective campaigns contribute to informed decision-making, public participation, and the overall credibility of the electoral process, aligning with the principles of SDG 16.

The observed moderate association between "*kampanye*" and "*pemilu*" reflects the attention given to election campaign within the analysed corpus. This association underscores the role of political communication and the exchange of ideas during elections. A well-conducted campaign can foster informed voter participation, contribute to the quality of political discourse, and enhance the legitimacy of electoral outcomes, which are essential aspects of achieving sustainable development goals.

The collocation of the term "*hasil*" (results) with "*pemilu*" (election) suggests a connection between discussions about election outcomes and the broader concept of elections within the analysed corpus. The observed collocate frequency of 23 instances in 14 texts, along with a Log Ratio (filtered) value of 1.907, indicates a notable but not extremely strong association between "*hasil*" and "*pemilu*."

In the context of the SDGs, the significance of election results is closely tied to SDG 16. Transparent, credible, and inclusive election results contribute to building strong democratic institutions and promoting accountable governance. Reliable election outcomes ensure that the voices and choices of citizens are accurately represented, fostering political stability and social cohesion.

The observed connection between “*hasil*” and “*pemilu*” highlights the attention given to election outcomes with the analysed corpus. It underscores the role of elections as a mechanism for citizens to express their preferences and influence the direction of governance. Election results influence policy decisions, leadership transitions, and overall societal development, all of which are relevant achieving the objectives of the SDGs.

The collocation of the term “DPT” (Final Voter List) with “*pemilu*” (election) points to discussion related to the voter list within the analysed corpus. The observed collocate frequency of 24 instances in 6 texts, along with a Log Ratio (filtered) value of 2.056, indicates a notable association between “DPT” and “*pemilu*.”

In the context of the SDGs, inclusive and transparent elections, supported by accurate and comprehensive voter lists like the “DPT,” contribute to SDG 16. The availability of a trustworthy and up-to-date voter list ensures that eligible citizens can exercise their right to vote without discrimination or exclusion. Moreover, it enhances the credibility of the electoral process, promoting a sense of fairness and legitimacy among citizens.

The observed connection between “DPT” and “*pemilu*” highlights the significance of accurate voter registration and the role it plays in ensuring a democratic and participatory electoral process. By facilitating the active engagement of all eligible citizens, including those from marginalized communities, a comprehensive “DPT” contributes to building inclusive and accountable governance systems that align with the principles of the SDGs.

The collocation “2019,” observed 66 times and surpassing the expected frequency of 7.083, suggests discussions about the previous 2019 General Election. Appearing in 18 texts, it indicates that discussions about the preceding election cycle continue to be relevant in the context of the 2024 General Election. This finding underscores the significance of building upon past electoral experiences, learning from successes and challenges, and ensuring the continuous improvement of the electoral process. The relation to SDG’s lies in the pursuit of inclusive and accountable governance (SDG 16), where effective electoral processes contribute to stable and responsive institutions that uphold democratic principles and human rights.

The collocation of the term “*sistem*” (system) with “*pemilu*” (election) indicates a moderate connection between discussion about election systems and the term “*pemilu*.” With the occurrences of 62 times in the corpus, 26 instances observed in 5 texts and a Log Ratio (filtered) value of 2.587, this suggests that considerations about the design, mechanisms, and fairness of the election process are relevant to the discourse surrounding the 2024 General Election. This finding holds implications for achieving SDGs 16, as a transparent and inclusive election system aligns with principles of good governance, fostering a just and accountable society. Therefore, attention to election systems reflects a commitment to SDG targets related to inclusive and responsible governance.

The high occurrence and frequency of the term “*penyelenggara*” (organizer) in relation to “*pemilu*” (election) with 57 times in the corpus and observed collocate frequency of 47 in 12 texts and a Log Ratio (filtered) value of 5.289 underscores a substantial association between discussion about election organizers and the term “*pemilu*.” This finding implies a crucial focus on the roles, responsibilities, and competence of organizers in ensuring a well-coordinated, transparent, and successful election process for the 2024 General Election. In the context of the SDGs, this strong connection aligns with SDG 16, emphasizing the importance of accountable and inclusive governance institutions for the enhancement of political processes and overall societal well-being.

The prominent observed frequency of the term “UU” (legislation) in relation to “*pemilu*” (election) with an observed collocate frequency of 47 in 11 texts and a Log Ratio (filtered) value of 5.804 underscores a robust linkage between discussions about legislative matters and the term “*pemilu*.” This finding suggests a significant emphasis on legal frameworks, regulations, and laws governing the electoral process for the 2024 General Election. This close association resonates with SDG 16, particularly in terms of promoting effective, accountable, and transparent institutions, which is pivotal for fostering inclusive and participatory democratic practices and upholding the rule of law for a more just and peaceful society.

Furthermore, unlike previous studies that have employed corpus linguistics to retrospectively analyze language, this research captures real-time language patterns, offering timely insights into the discourse surrounding the impending election. The emphasis on the Sustainable Development Goals (SDGs), particularly SDG 16, aligns this study with a broader global agenda, illustrating the interconnectedness of linguistic choices with principles of inclusive and accountable governance.

## 5. Conclusions

This study has successfully fulfilled its aim of uncovering linguistic nuances within the context of the 2024 General Election in Indonesia and evaluating their alignment with the United Nations' Sustainable Development Goals (SDGs). The analysis of collocates surrounding the term "*pemilu*" (general election) has revealed a rich tapestry of linguistic patterns, emphasizing the pivotal role of the upcoming election year, the structured stages of the electoral process, the inclusion of diverse participants, the significance of political campaigns, the impact of election results, the importance of the final voter list, and the commitment to transparent electoral systems and governance. These findings collectively underscore Indonesia's dedication to democratic values and contribute to the broader objectives of SDG 16, focusing on inclusive, accountable, and just societies. The nuanced examination of language usage in the electoral discourse provides valuable insights for political communication, governance practices, and informed decision-making, ultimately contributing to the nation's journey towards sustainable development.

It is important to acknowledge the limitations of this study. One limitation is the focus solely on linguistic analysis, which provides insights into language usage but may not fully capture the broader socio-political context of the 2024 General Election. Additionally, the study relies on corpus data from specific sources which might not represent the entirety of language usage surrounding the election. Furthermore, the analysis does not delve into the specific motivations or intentions, the findings contribute to a deeper understanding of the language associated with the 2024 General Election and its alignment with the principles of sustainable development.

## References

- [1] Artha, A. F. (2018). Revolusi pemerintahan, sudahkah berevolusi? Kolokasi adjektiva kata "Indonesia" dalam Coa dan Coha pada periode pemerintahan orde lama, orde baru dan era reformasi. *Etnolinguist*, vol. 2(1), 55–71. <https://doi.org/10.20473/etno.v2i1.8441>
- [2] Baker, P. (2010). Corpus methods in linguistics. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 93–116). Continuum International Publishing Group.
- [3] Crawford, W. J., & Csomay, E. (2019). Doing corpus linguistics. In *Journal of Chemical Information and Modeling* (Vol. 15, Issue 2).
- [4] Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins Publishing Company.
- [5] Laliy Zen, E. (2020). Corpus-driven analysis on the language of children's literature. *Icri* 2018, 17–22. <https://doi.org/10.5220/0009912700170022>
- [6] McEnery, T., & Hardie, A. (2011). *Corpus linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- [7] Oktavianti, I. N., & Ardianti, N. R. (2019). A corpus-based analysis of verbs in news section of the Jakarta Post: How frequency is related to text characteristics. *JOALL (Journal of Applied Linguistics & Literature)*, vol. 4(2), 203–214. <https://doi.org/10.33369/joall.v4i2.7623>
- [8] Purwitarini, D. (2020). The Words "Suspect" and "Patient": A Corpus-Based Analysis of Semantic Prosody in COVID-19 Corpora. *OKARA: Jurnal Bahasa Dan Sastra*, 14(2), 209. <https://doi.org/10.19105/ojbs.v14i2.3846>
- [9] Ramayanti, N. P. D., Sukarini, N. W., & Savitri, P. W. (2022). A Corpus Based Analysis of Adjectives Adorable. *PUSTAKA*, XXII(1), 55–58. <https://www.english-corpora.org/coca/>
- [10] Rosa, R. N., Amri, Z., and Zainil, Y. (2020). Translation strategies used by student translators in solving equivalence finding-related problems. *Advances in Social Science, Education and Humanities Research*, vol. 411, pp. 394-401. <https://doi.org/10.2991/assehr.k.200306.066>
- [11] Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Wiley-Blackwell.
- [12] Sudaryanto. (2015). *Metode dan aneka teknik analisis bahasa: Pengantar penelitian wahana kebudayaan secara linguistis*. Sanata Dharma University Press.
- [13] Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (p. 224). John Benjamins.
- [14] Yuliawati, S., Suhardjanto, T., & Hidayat, R. S. (2018). A corpus-based analysis of the terminology of the social sciences and humanities. *IOP Conference Series: Earth and Environmental Science*, vol. 175(1). <https://doi.org/10.1088/1755-1315/175/1/012109>