

The Image of "Nation" and "People" in the Indonesianwac Corpus: A Corpus Linguistic Study

Robi'atul Muthmainnah*, Tsanaa Mahara Haq, Mega Permatasari, Hikmah Aulia

Universitas Padjadjaran, Indonesia

Email: robiatul22001@mail.unpad.ac.id*, tsanaa22001@mail.unpad.ac.id,

mega22001@mail.unpad.ac.id, hikmah22002@mail.unpad.ac.id

ABSTRACT

The terms *bangsa* (nation) and *rakyat* (people) play a pivotal role as pillars of collective identity while simultaneously establishing political legitimacy. Both terms frequently appear in Indonesian socio-political discourse; however, the ideological nuances distinguishing their representations have not been extensively explored in depth. This study aims to reveal how the images of *bangsa* and *rakyat* are constructed in Indonesian discourse. The study employs a mixed quantitative and qualitative method using a corpus linguistics approach. This research utilizes data sourced from the IndonesianWaC corpus, accessed via SketchEngine, with the analysis limited to the top 50 collocates (determined by frequency and LogDice scores) as well as 50 relevant concordance lines for each lemma. The analysis focuses on three aspects: collocation patterns, observing lexical meaning associations; semantic prosody, identifying evaluative tendencies (positive or negative); and semantic preference, to map the semantic domains accompanying both words. This study is expected to provide empirical evidence regarding the dynamics of lexical usage that reflect power relations, societal attitudes, and identity construction in the Indonesian language.

Keywords: *prosody; preference; image; collocate; corpus.*

This article is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) 

INTRODUCTION

Politics in Indonesian discourse certainly recognizes the *nation* (*bangsa*) and the *people* (*rakyat*) as the main pillars that form a collective identity (Arifin et al., 2025; Azhar et al., 2024; Maryufani, 2022). A *nation* is understood as a broader cultural and historical entity, which emphasizes the commonality of shared origins, language, history, and cultural values. In line with what Sugono (2008) emphasized, a *nation* is a communion of individuals who have common ancestors, traditions, languages, and historical journeys, and have their own government. Meanwhile, the *people* include all citizens regardless of ethnicity or cultural background, who have the same rights and obligations in the eyes of the law.

Sugono (2008) also added that the *people* can be interpreted as the entire population of a country, civilians who interact with the government, ordinary people, and ordinary people in general. These two terms are not just identity labels but play an important role in the formation of political legitimacy, ideology, and social representation. Leaders use the *nation's* rhetoric more to strengthen unity and gain public support. On the other hand, the *people* are used to emphasize sovereignty and participation in government.

Political ideologies in Indonesia also use these two terms to formulate a vision of how the state should be run and who should be the holder of power. In the realm of public communication and socio-political discourse, the rhetorical function of the *nation* and the *people* is very significant. These two terms are widely used to arouse emotions, rally mass support, or criticize policies that are considered contrary to national interests and people's

aspirations. The use of these two terms can vary greatly depending on the context and the communicator's goals, but they both have the power to shape public opinion and influence policy direction.

Language plays a central role in shaping social representation. According to Chaer (2014:58), language as a means of human communication is human, which means that only humans have the ability to speak language and use it. This makes language not only a means of conveying information but also a foundation for the construction of social meaning and identity of individuals and groups. The implication is that each choice of words used contains social meaning, identity construction, and certain implicit interests. Lexical choices in discourse are important indicators of attitudes, values, and perspectives embraced by individuals or groups. Words that are consciously or unconsciously chosen reflect how a person or a group perceives the world around them.

Understanding this lexical choice makes it possible to unravel the complexity of the meaning contained in each utterance or writing. In the study of linguistics, understanding lexical meaning is essential to uncover how language functions structurally and functionally, as well as how the meaning of a word can affect the process of communication. Lexical meaning is also the basis for explaining various linguistic phenomena, ranging from the formation of metaphors to the shifts in meaning that arise throughout the development of a language, as explained by Putri (2024).

Therefore, lexical studies are very important in understanding the dynamics of power relations and ideologization in society. Through an in-depth analysis of words used in various contexts, it can be revealed how power works through language, as well as how certain ideologies are propagated and perpetuated through the use of language. Thus, lexical studies are used to have a more critical understanding of the social reality that exists around them.

Understanding how language shapes views about the *nation* and *people* is key to understanding the social and political dynamics in Indonesia. However, previous studies on lexical representation in the Indonesian language show some limitations that need to be overcome. There is a lack of research that explicitly compares the image of the *nation* and the *people* as different social and political entities. Most studies tend to discuss these two terms in general without dissecting the nuances of meaning and the ideological implications contained in them. In-depth research on the use of the words *nation* and *people* shows the need for an approach that can look at the patterns of use of the two terms more broadly, objectively, and based on evidence.

Until now, there have not been many studies that have explored the tendency of the meaning of the two words through the analysis of language use on a large scale. This approach is important because it can reveal how the *nation* and the *people* are used in various situations and contexts in society. Therefore, empirical evidence is quite crucial in understanding how the lexical representation of the two words is formed, as well as how their use can affect perceptions, attitudes, and social constructions in society. Further research needs to make more use of larger data, as Zhang (2022) said that large-scale corpus provides new perspectives and new ways to distinguish the meaning of word synonyms through natural context. Therefore, large-scale corpus is suitable for uncovering more deeply how "*nation*" (*bangsa*) and "*people*" (*rakyat*) are represented in Indonesian, as well as how these representations contribute to the construction of identity and social dynamics in Indonesia.

The Corpus Linguistic approach is one of the strong methodological foundations for understanding the empirical use of language in the study of lexical representation using big data. The analysis of the frequency, collocation, and lexical distribution patterns contained in this approach allows researchers to uncover relationships between language units that are not always apparent through intuitive observation. McEnery and Hardie (2012) revealed that this field is different from linguistics because in this field language-related research focuses on a series of procedures and methods for learning a language.

According to Cheng (2012), corpus linguistics is a field or linguistic approach that compiles, analyzes, and interprets language patterns based on empirical evidence in a machine-readable corpus. In corpus linguistics, the analysis of language data, both oral and written, is carried out using computer devices. In this way, language research can be done through the observation of large amounts of data more efficiently and objectively.

In addition, Rasyid, et al. (2022) stated that collocation analysis in accordance with theory has the purpose of answering all questions related to the nuances of related words. This in corpus linguistics can be referred to as *semantic prosody* and *semantic preference*. The concept of *semantic prosody* itself was originally proposed by Louw (1993). A word or phrase is said to have a negative or positive *semantic prosody* if it usually appears together with a unit that has a negative or positive meaning (McEnery and Hardie, 2012:136).

Then the concept of *semantic preference* was conveyed by Stubbs (2001) as a relationship, not between individual words, but between a lemma or word form and a set of words that are semantically related or can be said to be a field of meaning. Through this approach, lexical studies are more measurable, systematic, and scientifically accountable.

In the context of Indonesian research, the IndonesianWAC Corpus is present as a representative and relevant data source to observe the dynamics of lexical use in contemporary discourse. This corpus can be accessed through SketchEngine, one of the software for conducting corpus linguistics research. This 100 million-word corpus brings together various types of texts from various domains of discourse, such as news, opinion articles, institutional pages, and other online media, providing a rich context for analyzing the construction of meaning in public spaces. With a wide range of data, IndonesianWAC allows researchers to observe how certain terms are used, combined, and reproduced in various linguistic situations.

Of course, this research focuses on three main aspects, namely the pattern of collocation that forms the association of meanings, *semantic prosody* that shows evaluative tendencies, and *semantic preference* that reveals the tendency of the context or domain of meaning that accompanies the two words. Through the analysis of these three aspects, the research seeks to uncover how the image of the *nation* and *people* is formed, repeated, and interpreted in Indonesian discourse, as well as how these constructions reflect the social relations and ideological perspectives present in the use of language.

METHOD

This study implemented a systematic mixed-methods approach, starting from a quantitative framework for massive data collection and ending with an in-depth qualitative interpretation. The selection of this combined method—or *mixing methods*—was in accordance with what Brannen (2017) stated in his book, namely that the use of combined methods can be

said to be a way to examine the same research problem and thus increase the claims regarding the validity of conclusions that can be reached regarding the data.

The initial quantitative stage involved the use of Sketch Engine in the Indonesian Web Corpus (IndonesianWAC). The process began by entering the node words "*nation*" (*bangsa*) and "*people*" (*rakyat*) to find their basic concordances and frequencies. Collocation data were extracted with a span constraint of 4 (four words before and after the node) to obtain a list of adjacent words.

To ensure lexical accuracy, the extracted collocates were then statistically filtered using LogDice values. The use of LogDice was crucial because it measures the significance of collocational bonds; only the collocates with the highest LogDice scores, which show strong and non-coincidental relationships, were retained for further analysis.

The filtered quantitative data then entered a qualitative stage that focused on interpreting meaning. First, *semantic preference* analysis was carried out by grouping significant collocates into specific semantic categories (e.g., based on the UCREL system). This grouping aimed to uncover what domains of meaning or types of topics were consistently associated with "*nation*" and "*people*".

Next, the focus shifted to *semantic prosody* analysis to determine the affective imagery of the two node words. Each collocate was judged based on its sentimental value, whether positive (+), negative (-), or neutral (0). This assessment was qualitative and contextual, as the researcher had to refer to the concordance lines to ascertain the true affective meaning in the sentence.

The final stage returned to the quantitative realm to summarize the qualitative findings. The total frequency of positive, negative, and neutral collocates was calculated and converted into a percentage. The highest percentage of these affective categories definitively determined the dominant *semantic prosody* of the words "*nation*" and "*people*". Through tight integration between the statistical measurement of the corpus (LogDice) and the interpretation of the lexical context (*semantic analysis*), this study was able to deduce the image embedded in the two words.

RESULTS AND DISCUSSION

Analysis of the Collocation of the Word of the Nation

The words *nation* and *people* have similar meanings, but they both contain different nuances. As stated by Yuliawati (2018) that in the symbolization of a meaning, words not only have that referent, but certain connotations are also included in constructing, defending, or opposing the meaning itself. In its use, the word *nation* tends to give a more positive and authoritative interpretation, often associated with noble, majestic, and ideal things. Therefore, this word is more widely used to show a strong common identity and noble ideals that are visionary.

In essence, the word *nation* has various layers of meaning that make it semantically rich. Etymologically, the word *nation* is a translation of the English term *nation*, which is rooted in the Latin *natio*, which means "who has been born" or "born", and can be interpreted as descendant. Based on this basic meaning, the word *nation* refers to a group of humans, animals, or plants that have the same origin, characteristics, and distinctive traits. In other words,

everything that is born comes from a certain lineage that forms a group with its own history and development.

In the development of its meaning, the word *nation* then acquires a more specific and authentic connotation, especially in social and political contexts. This meaning refers to a group of people who have a common ancestral origin, customs, language, and history, and at a certain stage can have their own government. In addition, *the nation* also carries a nuance of values, namely a noble position or dignity, so that this term not only marks a group of origins, but also contains a spirit of togetherness, collective identity, and high ideals.

As researchers, we try to show that the word *nation* has certain nuances of meaning in its use. To do this, we took advantage of SketchEngine and analyzed it using LogDice values. Through this analysis, a number of collocates were obtained that represent the following tendency to use the word *nation*.

Based on the 50 collocates that have been found in the corpus data, the word *nation* has a strong association with the concept of national identity, unity, historical struggle, and development towards progress and prosperity. Categorically, the nation's collocate is dominated by nouns, such as unity, progress, struggle, culture, and welfare. The dominance of this noun shows that the word *nation* is generally present in constructions that describe the state, character, or dynamics of a nation, other collocates in the form of adjectives and verbs appear in smaller quantities, serving as an explanation of values (adjectives) or depictions of the collective actions of the nation (verbs). These findings confirm that the use of the word *nation* in Indonesian discourse tends to be conceptual and descriptive, closely related to issues of identity, history, and national development. Then, all of the collocate was analyzed in more depth through concordance of the word intellectually colloquial as exemplified as follows.

Based on the collocate, the word *nation* is associated with the word *educate* in the context of its use obtained through analysis in SketchEngine. One example that is clear is the emergence of cocate as stated in the 1945 Constitution, namely *public welfare, educating the nation's life and participating in implementing a world order based on it*. The use of the word *intelligent* is used in the context of the mandate or goals that must be achieved by a nation, especially in the field of education and improving the quality of human resources.

Overall, intelligent collocate is dominated by nominal nouns or phrases that function as the subject of the perpetrator, while right collocation is dominated by nominal phrases that are centered on the word *nation* as an object. This happens because *educating* is a transitive verb that requires a subject and object, and semantically is closely related to the goal of improving the quality of life of the nation as reflected in the discourse of education and national policies.

Based on the analysis of collocation, the word *nation* and *intellectualism* have a strong semantic relationship because *intellectualizing* as a transitive verb always places *the nation* as the object of purpose. The dominance of phrases such as *educating the life of the nation* shows that *the nation* is positioned as a collective entity that must improve its intellectual, moral, and social qualities. Thus, the relationship between the two illustrates the state's mandate in improving the quality of human resources while emphasizing the function of *educating* as a development action directed at the word *nation*.

Table 1. emphasizing

can maintain integrity and unity	Nation	and the state. For this reason, it is very necessary in the community
in order to maintain the integrity of the territory and the negative impact it has on integrity	nation.	In addition, straightening public opinion, and
Security Intrusions That Threaten Integrity	Nation,	territorial integrity and sovereignty of the nation and
Integrity of the Nation, Territorial Integrity and Sovereignty	Nation	and the state, the TNI always prepares itself
The hope of the nation depends on the integrity of the from threats and disturbances to integrity	Nation	Malay. Therefore, this convention has been
abroad. Maintaining integrity	Nation	and the State" (Law no. 34 of 2004 concerning the TNI)
Ummah, National Civilization, and Integrity	Nation	and play an active role in the life of the nation and
debt debt and the decline in the value of the rupiah, disintegration	Nation	need to be worked on to rise together
	Nation	threatens the integrity of our country, division

Based on the collocate that appears, the word *nation* is associated with the word *integral* in the context of its use obtained through analysis in SketchEngine. One example that seems clear is the emergence of cocate such as *security disturbances that threaten the integrity of the nation, territorial integrity, and national sovereignty*. The use of the word *integrity* in this context is used to emphasize the importance of a stable and undivided nation, and having territorial unity, identity, and sovereignty that is maintained.

Overall, the cocate of *integrity* is dominated by nominal nouns or phrases related to the concepts of defense, territory, and sovereignty, which function as objects or aspects that must be maintained, such as *territorial integrity, national integrity, and state integrity*. The right collocation of the word *integrity* is generally centered on the word *nation* or *state*, which indicates a collective entity that is the target of protection. This pattern arises because semantic integrity is closely related to stability, unity, and efforts to maintain national identity and integrity from various threats. This is reflected in the political, security, and statehood discourse that places integrity as an important element for the sustainability of a nation.

Based on the analysis of collocation, the words *nation* and *integrity* have a strong semantic relationship because integrity is used to mark the condition of the nation that must be guarded from the threat of division, disintegration, and security disturbances. The dominance of phrases such as *nation-building, territorial integrity, or state-building* indicates that *the nation* is positioned as a collective entity that needs protection in order to remain stable, united, and sovereign. Thus, the relationship between the two describes efforts to maintain national integrity while affirming the function of *integrity* as a concept of defense and unity directed at the word *nation*.

Collectively, the use of the word *nation* is dominated by the context that places it as an entity that is always in the process of formation and advancement. The nation is constructed as something that must continue to strengthen its moral, intellectual, and social qualities, while being protected from various threats that have the potential to damage its identity and integrity. This lexical tendency reinforces the image that the nation in discourse is an object that needs to be built and protected to remain strong and empowered.

Analysis of the Collocation of Folk Words

If the word *nation* tends to give a more noble, grand, and ideal interpretation, often associated with collective identity and visionary ideals, then *the word people* actually draws

attention to a more concrete, structural, and participatory reality. In its use, the *word folk* focuses on the led population, the subject of power, and the social base that is the daily spotlight of development and protection, bringing a sense of urgency to practical issues such as welfare and justice.

Etymologically, *the word folk* is absorbed from the Arabic language, *ra'iyyah*, which literally means a shepherd or flock under the protection of a leader. This basic meaning places the people as a group of people who are subjects or members of a state that are managed and protected. Based on this traditional meaning, *the people* refer to the population that is under authority, emphasizing their social and political functions as citizens bound to a specific governmental structure, and positioning them as entities that receive policy.

However, with the development of time with the entry of modern democratic ideology, the word *rakyat* has undergone a significant shift in meaning. *The people* are no longer merely *'the ones who are led,'* but have been transformed into the owners of supreme sovereignty and the main source of power. Concepts such as *people's sovereignty* affirm their active role and highest status as the legitimate owners of the state. In addition, the word *folk* is also widely used to refer to the masses or lower-stratified society (such as "the petty people" or "the poor"), which carries a strong connotation to issues of welfare and social justice, making it the center of attention in public policy.

In order to map the lexical image built around *folk* words in public discourse, researchers conducted in-depth research on their usage patterns. Through careful analysis of concordance, we managed to identify two dominant cocate characteristics, namely suffering and aspiration. These collocates serve as the main lexical markers that define the position of *the people's* subjects, and represent the tendency to use *the word folk* in social and political contexts.

Here are some sentences that demonstrate how *the word folk* collocates with *aspirations* in the text. This use consistently highlights the problem of political representation, when *the aspirations of the people*, including the demands of the oppressed, and important issues such as the eradication of corruption are often ignored, suppressed, or not represented by councillors and elites.

Based on the collocate that appears, *the word folk* is strongly associated with the verb *suffer* which places it as the object of the sufferer. This pattern shows the context of use obtained through the identification of the source of suffering. One example of a pattern that seems clear is phrases such as making our people suffer, making people miserable and continuing to suffer, or small people suffering more. The use of the word *suffering* is used in contexts that reflect the negative impact or adverse consequences of an event or policy (scarcity, price increases, corruption, political policies, war).

Syntactically speaking, collocate *suffers* dominated by verbs that function as predicates for *the people*. For example, "the people suffer" or a transitive verb that makes *the people* the object of suffering in the following sentence "make the people suffer". The left collocation is dominated by nominal nouns or phrases such as "poor people" or "small people" that serve as subjects experiencing suffering. Semantically, this association is closely related to the image of powerlessness and victimhood due to structural failure or elite actions, which is reflected in the discourse of social and political criticism.

In addition, there are also several sentences that display the concordance of the people's words as the object of the sufferer of the verb transitive by the perpetrator such as corrupt

officials or price policies. Suffering collateral emerged to define the existential conditions of the people. The analysis shows that phrases such as "small people will indeed suffer" or "our people suffer" are at the heart of discourse that criticizes the sources of people's suffering such as scarcity, price increases, and war.

The word *rakyat* is also closely associated with aspiration nominees that place it as the owner of the vote or the source of demand. This pattern arises in the context of failures and demands in the representation system. An example that often appears is the noun phrase *of the people's aspirations* which is then followed by a verb that indicates action against these aspirations, such as "does not / cares less about the aspirations of the people", "represents the aspirations of the people", or "accommodates the aspirations of the people". The use *of the word aspiration* is used in the context of political mandates and representative obligations that must be fulfilled by officials or representatives of the people.

Syntactically speaking, aspiration collocate (which is based on *the word folk* as an explainer) is dominated by nominal nouns/phrases that function as objects of suffering from verbs that indicate rejection or neglect (e.g., *the House of Representatives does not care about [people's aspirations]*). This happens because aspiration is semantically an output of the *people*, which then requires transitive verbs that describe the actions of the elite against that output. This association reflects the theme of political communication failure and representation dysfunction, as reflected in the discourse of demands and criticism of the performance of the representative council.

Collectively, the use of *the word folk* is dominated by contexts that describe situations of opposition to power. *The people* are constructed as entities that are afflicted by the actions of the elite (corruption, bad policies) and at the same time, their voices are ignored by the same elite. This lexical tendency reinforces the image that *the people* in this discourse are symbols for victims of social and political injustice.

Semantic Preferences and Semantic Prosodies of the Word *of the Nation and the People*

To understand how the word *Nation* and *Folk* Constructed in the Indonesian discourse, the researcher tries to explore the relationship between the meanings of the two words in depth through semantic preferences and semantic prosody preferences. Through semantic preference, the words *Nation* and *Folk* are the focus of this research. The words were analyzed based on words associated with the same word environment (kolocate) to map the field of meaning that surrounds them. Meanwhile, semantic prosody is used to reveal evaluative tendencies or auras in the two words to judge that they have a positive, negative, or neutral aura. Thus, researchers can see word imagery *Nation* and *Folk* in discourse.

In the early stages, the meaning aspect of the word *Nation* and *Folk* dissected by determining its significant corollates through the SketchEngine corpus software. With such devices, data can be extracted precisely. Then, the significance of the cocate was determined based on frequency to measure the regularity of word occurrence, as well as the LogDice score to measure the strength of word associations through IndonesiaWac corpus data which has a total of more than 100 million words. The researchers limited the analysis to the 50 collocates and 50 lines of correlation that had the strongest association values with the two words. This restriction is done to ensure that the analysis focuses on the dominant pattern that best represents the image *Nation* and *Folk* in discourse.

Based on the data findings, the researcher took one of the correlation samples that describes the emergence of the word *nation* with the collocate word *build*, which is generally related to the cultural-historical aspects of a country.

Table 2. Aspects of a country

Passionate behavior builds Nation and country.	Pride
as if forgotten in building Nation and the country in time	
The Constitution is "to contain the desire of the Nation Indonesia to build the country	
building and owning civilization Nation the superior and noble	
is a necessity for Nation Indonesia to continue to build	
We are determined to build Nation A united, just Indonesia	
all of us in building Nation and our country becomes	
The sector has not focused on building Nation healthy.	The influence of globalization,
Culture to build dignity Nation ignored .	Read More
time.	This shows psychology Nation they haven't built , no

Based on the correlation above, the word *construct* associated with the word *nation* shows a semantic prosody dominated by a positive aura. This is because the word *build* is a verb that inherently contains the meaning of progressivity or change to achieve better conditions. The use of this collocate emphasizes the position of the *nation* as an active subject (*agent*) who has the power to determine the direction of its own history and culture.

This is in accordance with what Isam (2012) said that a lexicon can be said to have a positive value if the lexicon is in an atmosphere of collocation that is believed to be positive and vice versa. However, the use of cocate is not always accompanied by a positive aura. Some concordances are seen to indicate having a semantic prosody that has a negative aura because it is associated with a negative context. This can be seen from the correlation line: *culture to build the dignity of the nation is ignored*. In this data, the word *build* is juxtaposed with the passive verb *ignored* which has a negative meaning. Thus, the word *build* can also be used as a form of criticism or social concern.

Meanwhile, correlation that describes the emergence of *the word folk* with the colloquates of the word *poor* is closely related to the issue of social and economic disparities. In contrast to *the nation* that is imaged of glory, the frequency of *people's* words that appear in the discourse is dominated by financial vulnerability and powerlessness. This can be seen from the correlation below.

The data shows that the *word poor* in *folk* words shows a semantic process dominated by a negative aura. This proves that *the people* are placed in a subordinate position, namely as a vulnerable group that needs help or support from other parties. In the construction of this discourse, the people lost their role of control as the holder of the highest sovereignty and shifted to the role of being a victim of structural injustice (policies, infrastructure, and social practices). This is shown in the line of corcondance: *not taking sides with most of the poor who are in the lower structures*.

However, not all appearances of *the word poor* show an absolute negative aura. In accordance with what Bednarek (2008) said, negative or positive collocation can be an indication but not the basis for making a hypothesis but depends on the verb, even though the verb only appears in a negative context and even though its placement with non-negative expressions is not grammatically appropriate, the verb does not acquire a negative connotation.

In some collocates, *the word poor* is present in a more positive semantic process because it is juxtaposed with benefactive nouns or phrases that support the fulfillment of basic rights. This can be seen in the correlation: *free medical treatment for the poor, free education for the poor*. In this context, even though *the word folk* is labeled as the word *poor*, the meaning of the focus of the discourse shifts from "inequality" to "solutions" or "social security" provided by the state, so that the image that is built is an effort to improve.

From the two examples of analysis above, it can be concluded that there is a strong dichotomy pattern between representations *Nation* and *Folk*. This conclusion is reinforced by the distribution of quantitative data of semantic prosody presented in the following diagram.

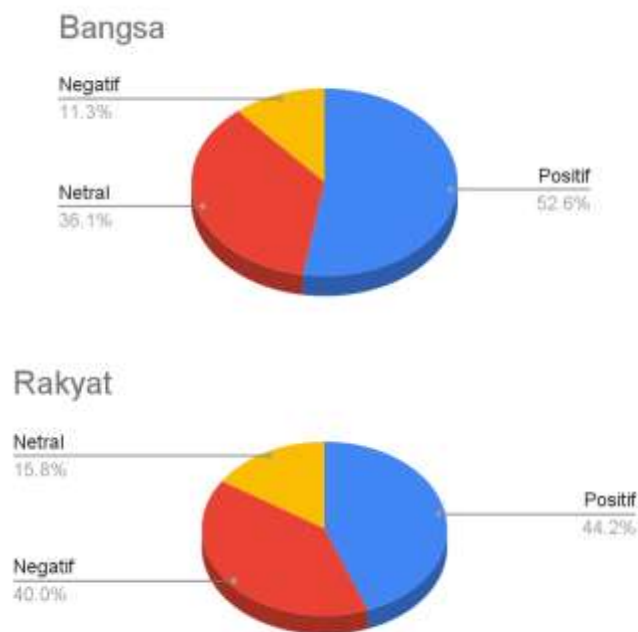


Figure 1. Semantic Prosodi Diagram of the words *nation* and *people*

The diagram above presents the results of the analysis calculation of 50 selected collocates and 50 concordance lines that have been classified into the categories of positive, negative and neutral. The data confirms a significant difference in evaluative value. The word *nation* has a dominant positive prosody which reaches 52.6% with a very minimal negative percentage, which is only 11.3%.

This figure indicates that in public cognition and discourse, *the nation* is a noble, proud, and optimistic entity. This condition is inversely proportional to the people's words. The positive prosody, said *the people*, recorded 44.2%, the positive value was mostly related to the context of receiving assistance or support received by the people to get their rights. In addition, the percentage of negative prosodies jumped dramatically to reach 40.0%. The negative figure shows a surge four times greater than the negative prosody, said *the nation*. This shows that *the people* are the entities most often associated with suffering, misery, and inequality.

This image inequality is further clarified through semantic preference. This study classifies the collocation that appears in the word *Nation* and *Folk* into several specific categories of meaning.

Of the 14 categories of semantic preferences identified in the IndonesianWaC corpus, the word *nation* is: 1) social integrity; 2) Excellence; 3) time (historical); 4) empowerment; 5) change; 6) protection; and 7) honor; 8) movement; 9) social groups; 10) religious and supernatural; 11) the things of life; 12) art or creativity; 13) obstacles; 14) Character. From the results of this analysis, the use of the word *nation* is closely associated with a high "sacred" and "historical" collective entity.

This can be seen from the categories of social integrity, excellence, time, honor, movement, social group, art or creativity, character, religion and the supernatural. In addition, the word *nation* is also associated with the narrative of "progress", such as in the category of empowerment and change to lead to better conditions. Then, *the nation* is also used as a powerful entity, which can be seen from the categories of excellence and protection. However, the word *nation* can also indicate aspects of obstacles (such as problems) in the process of its dynamics. Thus, it can be said that the word *nation* reflects a national identity that is loaded with integrity and honor, as well as presenting an entity that is resilient and able to survive to defend its sovereignty.

In contrast to the nation's construction *which* tends to be idealistic, the semantic preference of the *people's* words is inversely proportional to realistic. Here is a table of semantic preferences on *folk* words categorized into several categories based on UCREL:

Based on the 13 categories above, it indicates that the semantic preferences of *the people's* words are: 1) gap; 2) power relations; 3) barriers; 4) Excellence; 5) hard work; 6) negative emotions; 7) politics; 8) change; 9) employment; 10) religious and supernatural; 11) general/ordinary; 12) law; and 13) aspiration and communication.

From the results of this analysis, the use of *the word folk* is closely associated with structurally "vulnerable" and "depressed" entities, such as inequality, negative emotions, employment, and power relations. Meanwhile, the word *people* is associated with "fighting power" or resistance, such as hard work, obstacles, politics, law, as well as aspirations and communication that lead to demands for justice.

This can be seen from the aspect of strong aspirations and politics as a form of response to pressure. Even so, the word *rakyat* can also show an aspect of superiority (sovereignty), but this aspect is often present along with the narrative of the struggle against obstacles. There are colloquates that show superiority, but these aspects are often present along with the narrative that is still being fought. It can be said that the word *folk* reflects the reality of social inequality, the burden of suffering, and political entities that continue to move to voice rights.

From the results of this analysis, it can be concluded that there is a fundamental difference in the development of the meaning field between the words *nation* and *people*. The field of national meaning consistently leads to idealism, strength, and honor, while the dominant field of meaning in *the people* is related to vulnerability, material shortcomings, and resistance.

Regarding the field of meaning seen from its semantic preferences, it can be concluded that *the noun nation* is more widely used to discuss the state as a great abstract concept. On the other hand, the *people's noun* is more widely used to discuss the reality of civil society with all its burdens of suffering and political dynamics. This can be seen from the accompanying words such as *building* and *dignified* for *the nation*, as well as *poor* and *commoners* for *the people*. Thus, it can be concluded that in the discourse of IndonesianWaC, *the nation* is present as a

symbol of legitimacy and pride, while *the people* are present as a representation of the reality of social inequality that still must be fought.

Comparison of Analytical Implications in Public Discourse

Based on the analysis of the data that has been carried out, it can be said that the findings of the collocation show that the word *nation* tends to be associated with words that are nuanced, institutional, and of high value. This is clearly shown in the collocates of *intellectual*, *welfare*, *wholeness*, and so on so that the semantic preferences produced are social integrity and honor.

The semantic preference of the word *nation* signifies that the concept of the word is positioned as a word to denote one that represents an ideal entity with noble norms. Then, based on the semantic prosody found through analysis, the word *nation* has a positive domain. This is shown by the existing percentage, namely a positive aura of 52.6% which dominates the 50 data collates found.

In contrast to *the word nation*, the word *people* is more often associated with a collocation that revolves around things that indicate the economy, inequality, social, and others. This is also clearly shown by the semantic preferences of the reality of social inequality, political entities, and resistance. Even so, the semantic prosody is enough to show an almost balanced percentage, namely 44.2% positive aura and 40.0% negative aura.

The positive concept in the people's word is related to the collateral of *welfare*, *sovereignty*, *prosperity*, and so on. Meanwhile, negative concepts contained in *folk* words usually come with problematic issues or collocates that show more opposition such as miserable collate, *suffering*, *oppressive*, *poor*, *commoner*, and so on. This shows that *the word folk* is associated with things that are more concrete and related to social reality and are more often used as something that is not good than *the nation*.

In contrast to the reality that we have known so far, the *people* always have a bad connotation related to oppression. However, in the results mentioned, it was found that *the word folk* is only part of the social political discourse to show reality without dreams or high morals. Belief related to *the word folk* that has a connotation or bad aura is conveyed by Suwignyo (2020) who states that the word *folk* has a narrowing of meaning around the 1930-1970s which initiates that the word has an unacceptable meaning in the midst of the new regime in the world of Indonesian politics and brings threatening consequences and causes "left-left" associations. Therefore, the word *nation* is more often used to show hope and noble morals in formal political discourse such as in the Youth Pledge, Proclamation of Independence, and so on.

It can be said that the comparison of these two words indicates that the word *nation* carries the image of a more ideal country with certain noble norms, while the word *people* favors a more realistic image with social circumstances with more diverse emotional dynamics than the word *nation*. Then this causes a certain word framing, word framing is also related to word selection and this is in accordance with what was conveyed by Rahma (2025) that word selection reflects a social and cultural view related to a certain society or association according to the place where the word is used.

which is related to the words *nation* and *people* which usually occur in media discourse and also the public. Usually, the word *nation* is used in more formal media discourse such as a

government statement, policy document, and prioritizing news with pride and representation. The word *nation* is a tool to represent a good ideology towards a country.

Meanwhile, the word *people* is usually used in the context of political discourse that is more about social coverage and news about a policy that is quite detrimental to the people. The discourse used in this public space positions the word *people* as a figure who becomes a victim, feels the consequences, or becomes the basis for the legitimacy of political actions. This difference in *framing* or framing of words shows that *the nation* is used to build a certain morality while the word *people* is more often used to negotiate an interest or related to the discourse of social reality.

These findings provide theoretical implications for lexical studies and discourse analysis. Lexically, research proves that the differences in prosody and semantic preferences between *the nation* and *the people* are not only the result of differences in the meaning of dictionaries, but also the product of usage patterns in public discourse. This supports the argument that the meaning of words is shaped by the context in which they are used in society (*usage-based semantics*). Discourse-speaking, the use of these two words shows that the language of the media plays a central role in building social and political image. *The nation* produces identity, while *the people* produce power relations and legitimacy. These findings can contribute to the study of political representation, linguistic ideology, and the process of constructing meaning in media.

By combining the results of the analysis of collocation, semantic prosody, semantic preferences, and *discourse framing*, it can be concluded that the two terms perform different representational functions in the IndonesianWac corpus. *The nation* exists as an ideal entity that represents collective ideals while *the people* appear as a social entity that experiences concrete realities and often becomes the legitimacy of political actions. This integration confirms that lexical choices in discourse are not only linguistic aspects, but also ideological actions. Thus, this discussion presents a complete narrative of how language contributes to the production of social and political meaning in Indonesian media.

CONCLUSION

Analysis of the lexical construction of the words *nation* (*bangsa*) and *people* (*rakyat*) in the IndonesianWAC corpus reveals a sharp ideological dichotomy, as evidenced by 50 collocates and concordance lines that reflect underlying ideologies and power relations in language use. The *nation* emerges as a sacred, historical, and development-oriented entity, dominated by positive *semantic prosody* (52.6%) and *semantic preferences* linked to social integrity, excellence, and honor, positioning it as a legitimate state symbol and active subject in cultural, historical, and developmental narratives. In contrast, the *people* are framed as a vulnerable, structurally repressed entity, with negative *semantic prosody* nearly four times higher than that of *nation*, alongside *semantic preferences* highlighting gaps, power imbalances, and barriers—often depicting them as policy victims or rights-seeking strugglers rather than sovereign holders. This hierarchical discursive framing underscores distinct roles in Indonesian socio-political rhetoric. For future research, a diachronic corpus analysis is recommended to trace shifts in these meanings over time, enhancing contributions to lexical semantics and critical discourse analysis.

REFERENCES

- Arifin, F., Tedjabuwana, R., Wiyono, S., & Abdullah, M. K. bin. (2025). Indonesia's identity politics and populism: Disruption to national cohesion. *Jurnal Civics: Media Kajian Kewarganegaraan*, 22(1), 166–175.
- Azhar, U., Yaseen, S., Arif, M., Babar, M., & Sagir, M. (2024). Emission of greenhouse gases from transportation. In *Advances and Technology Development in Greenhouse Gases: Emission, Capture and Conversion* (pp. 147–163). Elsevier.
- Bednarek, M. (2008). Semantic Preference and Semantic Prosody Re-Examined. *Corpus Linguistics and Linguistics Theory*, 4, (2). 119-139. DOI 10.1515/CLLT.2008.006
- Brannen, J. (2017). Mixing Methods: qualitative and quantitative research. <https://doi.org/10.4324/9781315248813>
- Chaer, A. (2014). *Linguistik Umum*. Jakarta: Rineka Cipta.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. Routledge: Milton Park, Abingdon, Oxon / New York.
- Isam, H., & Awal, N. M. (2012). Nilai Setia Dari Perspektif Prosodi Semantik: Analisis Berbantu Data Korpus. *GEMA Online Journal of Language Studies*, 12(2).
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp.157-175). Amsterdam: John Benjamins Publishing Co.
- McEnery, Tony dan Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory, and Practice*. Cambridge University Press.
- Putri, N. S. (2024). Makna Leksikal: Pondasi Pemahaman Makna dalam Analisis Semantik. Diakses pada tanggal 22 November 2025, dari <https://www.jurnalissubar.com/2024/10/makna-leksikal-fondasi-pemahaman-bahasa.html>
- Rahma, N. N. 2025. Prosodi Semantik Sinonim Kata “Laki-Laki” Dan “Pria” Dalam Wacana-Wacana Di Internet: Studi Linguistik Korpus. *Paramasastra: Jurnal Ilmiah Bahasa Sastra dan Pembelajarannya*. Vol. 12 No. 2. <https://doi.org/10.26740/paramasastra.v12n2.p265-281>
- Rasyid, I. R., Laura, H. M., & Rachman, F. (2022). Semantik Prosodi Nomina Indonesia dalam Korpus Wacana Berita Arab Internasional Oman 2013. *Jurnal Al-Azhar Indonesia Seri Humaniora*, 7(3), 192-198.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Sugono, D. (2008). *Kamus Bahasa Indonesia*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.
- Suwignyo, A. (2020). Rekayasa Bahasa dan Konstruksi Politik “Persatuan-Kesatuan” dalam Wacana “Negara Kesatuan Republik Indonesia”. *Jurnal Kawistara*, 10, (1). 116-132.
- Yuliawati, S. 2018. Perempuan atau Wanita? Perbandingan Berbasis Korpus tentang Leksikon Berbasis Gender. *Paradigma, Jurnal Kajian Budaya*, 8(1), 53. <https://doi.org/10.17510/paradigma.v8i1.227>

Zhang, L. 2022. Studi Berbasis Korpus: Perbandingan Kolokasi dan Prosodi Semantik Sinonim Bahasa Indonesia “Menyebabkan” dan “Mengakibatkan.” *Mabasan*, 16(1), 153–176. <https://doi.org/10.26499/mab.v16i1.517>