



LEXICAL PATTERNS IN THE KAPAI-KAPAI MANUSCRIPT BY ARIFIN C. NOER BASED ON CORPUS THROUGH ANTCONC

Arif Hidayat, Silvia Al Viana

Nusa Mandiri University, Monash University

arif.aft@nusamandiri.ac.id¹, svia0003@student.monash.edu²

ABSTRACT

The Kapai-Kapai script by Arifin C. Noer occupies a significant position in modern Indonesian theater. It is a work known for its symbolic power and distinctive linguistic expression. The study of lexical patterns in this script is necessary to understand how word choice and connection shape discourse cohesion and support the creation of dramatic effects. This study aims to identify and interpret lexical patterns in Kapai-Kapai through a corpus approach with data processing using Antconc. The script text is converted to plain text format and analyzed using word list, concordance, and collocation features to map word frequency, context of occurrence, and lexical relationships. The results of this study show that lexical repetitions such as *abu*, *iyem*, and *kau* create a poetic rhythm and reinforce the image of destruction and mortality; thematic cohesion is formed through pairs such as *iyem-emak* and *kelam-bulan* that unite the interpersonal and atmospheric field; while symbolic collocations such as *cermin-pintu* and *mati-sampai* deepen the metaphorical landscape of the performance. This finding confirms that the lexical patterns in Kapai-Kapai function dually as cohesive devices and dramaturgical mechanisms that build intentionality while strengthening Arifin C. Noer's distinctive ideological message.

Keywords: lexical patterns, corpus, AntConc, theater script, Kapai-Kapai, Arifin C. Noer

INTRODUCTION

As one of the maestros of modern Indonesian theater, Arifin C. Noer is known for presenting scripts full of social, political, and psychological complexity embodied through distinctive language constructions. Kapai Kapai, one of his famous works, represents various forms of social class opposition triggered by social relations factors and intercultural misunderstandings (Fikri & Hasanuddin WS, 2023). In the Mega-Mega script, illocutionary acts are used to show the complex interaction of characters and ideologies (materialism and humanism) to reflect the hegemony of the dominant group over the subordinate group (Ahmad & Herdiana, 2022; Safira et al., 2024). Meanwhile, in *Sumur Tanpa Dasar*, psychological conflicts influenced by the clash between social expectations and personal desires are emphasized, thus complicating the dramatic language constructed (Turahmat, 2022). These various representations show that language in Arifin's works functions as a medium of communication and a dramaturgical device that binds ideas and emotions.

To uncover dramaturgical devices in linguistic aspects, it is necessary to explore the uniqueness of Arifin C Noer's language through lexical patterns. Lexical patterns refer to the arrangement of words or terms that frequently appear together in language use (Nogier & Zock, 1992; Sánchez et al., 2023). The study of lexical patterns allows for an understanding of the distribution and relationships of words in a text, both manually and through the application of computational linguistics such as natural language processing, machine translation, and semantic text analysis. Lexical patterns encompass the study of vocabulary, meaning, and its relations in a language, including the process of derivation, structure, and syntactic connotation (Sidorova et al., 2025). Functionally, lexical patterns become linguistic units that connect ideas and reality through meaning, formed by syntactic rules that give rise to relationships between words (Sánchez et al., 2023). The uniqueness of the distribution of lexical patterns in each language reflects the differences in how the language constructs meaningful relations. In this study, lexical patterns are used to reveal how language in Kapai-Kapai is organized as part of a dramaturgical device.

The corpus linguistic approach provides a systematic framework for analyzing language in theatrical scripts. A corpus of theatrical scripts can be a single play, a number of works by a single author, or a collection of scripts from a specific period and genre that have been digitized for linguistic analysis (Madieva et al., 2015). Theatre scripts have advantages as corpus data because they contain natural representations of spoken language, clear dialogue structures, and social contexts reflected through variations in register and style between characters. Using the Antconc application, lists of the most frequently occurring words and phrases can be generated to help identify themes, analyze collocations to see word associations that form relays between characters, and examine concordances to understand word usage in context (Sánchez et al., 2023).

Lexical pattern studies have been conducted in various genres, ranging from analysis of English verbs (Qiu, 2020), ESL student writing (Qin, 2023), lexical metaphors in news texts (Safitri et al., 2023), to lexical ambiguity in general contexts (Mukhlis, 2022). Although contributing to the understanding of lexical phenomena, these studies have not yet touched on the realm of Indonesian theater scripts, let alone examining the works of Arifin C. Noer, who holds a central position in the development of modern theater. There have also been no studies that utilize a corpus approach with Antconc tools to map lexical patterns in theater scripts. This situation opens up space for research to identify and classify lexical patterns in character formation in dramatic texts.

METHOD

This study uses a descriptive qualitative approach supported by corpus linguistic analysis. This type of corpus research was chosen to provide quantitative accuracy in measuring frequency, collocation, and lexical distribution. Therefore, the findings are empirically verified and replicable. The data source is the Kapai-Kapai manuscript by Arifin C. Noer. This manuscript was selected based on its historical and artistic significance in modern Indonesian theater. The digital Kapai-Kapai manuscript was converted to plain text format to facilitate processing in AntConc software. The data cleaning stage included removing stage directions irrelevant to lexical analysis, standardizing the spelling of character names, and correcting conversion errors. The analysis was conducted using AntConc's three main features: Wordlist, concordance, and collocation. Wordlist is used to identify dominant vocabulary and potential key themes. Concordance searches keywords to uncover lexical meanings that are appropriate to



dramatic usage. Collocation is used to uncover associations of meaning, symbolism, and relationships between characters. The findings were validated by manually cross-checking the output to avoid errors in reading the application.

FINDINGS

The lexical pattern analysis in this study began with the selection of target words based on the wordlist generated by AntConc software. The wordlist displays all vocabulary in Kapai Kapai along with their frequency of occurrence. This allows the intensity and potential thematic connections within the text to be identified. From the list, ten words were selected with varying frequency distributions and high rankings, such as *abu* (342 occurrences, rank 1), *iyem* (142 occurrences, rank 3), *kau* (107 occurrences, rank 4), and *menangis* (11 occurrences, rank 84). This selection is technical and refers to the position of the words in the wordlist rankings summarized in Table 1.

Table 1 ten target words for concordance and collocation analysis

No	Target Word	Frequency	Rank
1	abu	342	1
2	iyem	142	3
3	emak	76	7
4	Majikan	53	15
5	kelam	47	17
6	bulan	37	20
7	cermin	22	50
8	mati	13	76
9	Menangis	11	84
10	kau	107	4

After determining the ten target words from the wordlist, the next step is to examine the concordance of each word. Concordance analysis is conducted to examine the context in which the target words appear in the Kapai-kapai manuscript, both on the left (L) and right (R) of the word. This concordance data presents sentence fragments (concordance lines) that show the actual use of the words in the dialogue, allowing the syntagmatic patterns that accompany them to be identified.

Table 2 Concordance Results of Ten Target Words

No.	Left Context	Target Word	Right Context
1	penuh debu dan	abu	menempel di bajumu
2	warna langit jadi	abu	kehitaman sore itu
3	kata si	iyem	sambil menunduk
4	sudah bilang ke	iyem	bahwa majikan marah
5	aku rindu	emak	waktu kecil
6	tegur	majikan	dengan suara keras
7	suasana malam yang	kelam	menyelimuti panggung
8	cahaya	bulan	masuk ke jendela
9	tatap di	cermin	pecah itu
10	semua diam saat	mati	lampu panggung
11	ia mulai	menangis	tanpa suara
12	apa yang	kau	mau dariku

Table 2 above contains a list of concordance lines for the ten target words: *abu*, *iyem*, *emak*, *majikan*, *kelam*, *bulan*, *cermin*, *mati*, *menangis*, and *kau*. Each entry displays the target word in the center, surrounded by left and right contexts taken directly from the script. For example, the word "abu" frequently appears in word constructions involving greetings to other characters on the left, while "iyem" is often accompanied by verbs on the right. Meanwhile, the word "kelam" appears more frequently in mood settings, such as phrases containing time elements. This data provides an initial overview of the syntagmatic distribution of each target word.

The final stage of this corpus analysis was collocation mapping for the ten selected target words. Collocation indicates the tendency of two words to appear close together within a certain range. This analysis was conducted to identify lexical pairs that are highly related as an indication of the connection between meaning and dramatic function within the text.

Table 3 Collocation of ten selected target words

Target Word	Kolokat	Rank	Freq(L+R)	FreqL	FreqR	Likelihood
abu	majikan	1	117	61	56	159.27
iyem	emak	1	66	33	33	158.84
emak	iyem	1	66	33	33	158.84
majikan	abu	1	117	56	61	159.27
kelam	bulan	1	7	5	2	31.14
bulan	kelam	1	7	2	5	31.14
cermin	pintu	1	7	4	3	31.58
mati	sampai	1	7	4	3	30.11
menangis	jangan	1	8	6	2	28.45
kau	iyem	1	22	10	12	32.55

Table 3 above displays the most powerful collocation pairs for each target word. For example, the word "abu" has the primary colloquial "majikan" with a likelihood of 159.27 and a balanced distribution of occurrences on the left (61) and right (56) sides. The words "iyem" and "emak" are each other's top collocates, each appearing 66 times with a symmetrical distribution. In the word "kelam", the strongest collocation is "bulan" (likelihood 31.14), while the word "cermin" is closely associated with "pintu" (likelihood 31.58). The word "mati" tends to collocate with "sampai" (likelihood 30.11), while "menangis" has the primary pair "jangan" (likelihood 28.45). For the word "kau", the strongest collocation is found with "iyem" (likelihood 32.55).

DISCUSSION

An analysis of lexical patterns in Kapai-Kapai demonstrates how Arifin C. Noer's language is a dramaturgy that combines meaning and thematic symbolism. Findings from the wordlist, concordance, and collocation provide a comprehensive picture of how words function in the text. The target words play a crucial role in the plot, such as *iyem*, *emak*, *majikan* which leads to the relationship between characters, *kelam*, *bulan*, *cermin* which gives color to the atmosphere, as well *menangis*, *mati*, and *kau* which brings out the emotional dimension.

These ten words have a high frequency and play a strategic role in the construction of the dramatic text. In terms of lexical frequency, the words *abu* (342 occurrences), *iyem*



(142), and *kau* (107) occupy the top three positions. This high frequency is not random, but builds a planned repetition that binds the structure of the dialogue and creates a verbal rhythm. *Abu*, for example, serves as a marker for objects and becomes a central metaphor that recurs in various scenes to signify destruction and mortality. In the dialogue fragment "*Semua akan jadi abu, Iyem*" (Everything will be ashes, Iyem), *abu* is used as a metaphor for finality, signifying that everything will eventually be destroyed or destroyed.

Second, regarding lexical cohesion, the repeated occurrences of words *iyem* and *emak* (142 and 76 times respectively) indicate a network of cohesion built through kinship relations. This repetition creates thematic continuity, while synonym-thematic pairs such as *kelam* (47 times) and *bulan* (37 times) connects the image of darkness with the sense of time and the life cycle. This pattern aligns with Wu and Ajmal et al.'s view that lexical repetition strengthens discourse unity and makes it easier for the audience to follow the emotional flow of the text (Ajmal et al., 2022; Wu, 2024).

Third, within the framework of dramatic diction, the choice of words has symbolic nuances such as *cermin*, *mati*, and *menangis* not only serves a descriptive function but also forms an imaginative landscape for the performance. For example, *cermin* (22 times) acts as a visual object that reflects the character's inner condition, while *menangis* (11 times) is a direct signal at the peak of emotional intensity. In line with Dmytruk & Ostapenko, this type of diction stimulates the audience's emotional response and enriches the aesthetic experience (Dmytruk & Ostapenko, 2022).

Concordance of target words in *Kapai-Kapai* shows how the frequency of appearance combines with lexical cohesion to build the dramatic layer of the script. Said *abu* for example, often appears side by side with *majikan* or *iyem*, indicating that the image of destruction is often connected to the relationship between characters. Likewise *kelam* which appears with the word *bulan*, forming an atmospheric collocation that combines the visual image of night with a gloomy and fatalistic feel.

The words *iyem*, *emak*, and *kau* dominate interpersonal conversations, creating cohesion through repetition that emphasizes emotional or hierarchical relationships among characters. While words like *mati* and *menangis* has less frequent, but still strategic, occurrences, namely appearing at dramatic points to mark emotional transitions or the climax of tragedy. In this case, Arifin C. Noer's dramatic diction utilizes highly emotionally charged words not only as thematic markers, but also as triggers for the audiences' affective response, in line with the concept 'dramatic diction' which places vocabulary selection as the main rhetorical instrument (Dmytruk & Ostapenko, 2022).

The presence of these words in various parts of the manuscript creates a strong thematic cohesion: *abu*, *kelam*, *mati*, and *menangis* organizing the meaning field of destruction and mortality, whereas *iyem*, *emak*, and *kau* reinforces the emotionally charged interpersonal field. The interaction of these two fields makes the text not just a dialogue, but also a symbolic landscape that blends social and existential realities.

Collocations of target words reveal patterns of lexical relatedness that form a network of meaning within *Kapai-Kapai*. Words pair *abu-majikan* and *majikan-abu* signify the symbolic relationship between destruction (*abu*) and authority figures (*majikan*), which creates an ironic contrast: power stands on a fragile and ephemeral foundation. Likewise *iyem-emak* and *emak-iyem* show intimate personal bonds, where family relationships become one of the axes of dramatic interaction, often containing nuances of warmth as well as helplessness.

Collocation of *kelam-bulan* and *bulan-kelam* form a strong poetic image; *bulan* usually associated with the beauty of the night is transformed into a part of a gloomy

landscape, thus reinforcing Arifin C. Noer's characteristic atmosphere of absurdity. On the other hand, *cermin-pintu* opens the metaphorical layers: *cermin* as a self-reflection and *pintu* as a threshold, hinting at inner struggle and the possibility of existential transition.

Collocation of *mati-sampai* indicates the intensity of a dramatic situation, where death is positioned as the end point of the character's journey. Meanwhile *menangis-jangan* emphasizes the form of prohibition or emotional restraint, showing the dynamics of emotional control among characters. On the other hand, *kau-iyem* underlines the direct, personal relationships between characters, often in dialogue that is raw and confrontational.

This collocational connectedness not only strengthens lexical cohesion through association of meaning but also functions as 'dramatic diction' which guides the audience towards a symbolic reading. Words paired in collocations reinforce the field of meaning identified in the wordlist and concordance, creating a consistent web of symbols, relationships, and emotions throughout the text.

Overall, the analysis of wordlist, concordance, and collocation in *Kapai-Kapai* reveals four main fields of meaning that form the lexical pattern of the manuscript. First, the field of relationships among characters supported by repetition *iyem*, *emak*, *majikan*, and *kau*, marks personal interactions that are full of hierarchy, emotional closeness, or direct confrontation. Second, the field of atmosphere and atmosphere created by *kelam*, *bulan*, and *cermin*, presents a visual and symbolic landscape that combines darkness, time cycles, self-reflection, and existential transition. Third, the dramatic emotional field that emerges from *menangis* and *mati*, shows the points of emotional intensity and the climax of the tragedy. Fourth, the field of destruction and mortality centered on *abu* positions the metaphor of finality as a thematic knot that binds the entire narrative. The interaction among these fields forms a lexical cohesion that not only maintains the continuity of discourse but also strengthens Arifin C. Noer's 'dramatic diction' so that the text has double functions as a dialogue space and a symbolic landscape rich in meaning.

Conclusion

Lexical pattern analysis of the manuscript *Kapai-Kapai* created by Arifin C. Noer shows that vocabulary selection not only functions as a means of conveying literal meaning but also as a dramaturgical device that builds the thematic structure, atmosphere, and emotions of the performance. The frequency of certain words such as *abu*, *iyem*, and *kau* demonstrates repetition strategies that create the rhythm of dialogue while binding narrative elements. Concordance reveals the context in which words appear, emphasizing specific meanings—from relationships between characters, to gloomy moods, to symbols of destruction and mortality. Meanwhile, collocations reveal lexical connections that add symbolic depth, such as pairs *abu-majikan* which hints at the irony of fragile power, or *kelam-bulan* which changes the image of the beauty of the night into a gloomy landscape.

The four identified fields of meaning—character relationships, mood and atmosphere, dramatic emotion, and destruction and mortality—form a consistent network of lexical cohesion throughout the text. This finding confirms that dramatic diction in Arifin C. Noer's work operates on two levels simultaneously: conveying the story through dialogue and constructing a symbolic landscape that invites the audience to engage emotionally and reflectively. Thus, this study demonstrates that lexical pattern analysis



can be an effective entry point for understanding artistic strategies and symbolic power in modern Indonesian dramaturgy.

The findings of this study have important implications in three areas. First, theoretically, the analysis of lexical patterns in Kapai-Kapai enriches the study of lexical cohesion and dramatic diction in Indonesian dramaturgy, particularly by showing how word frequency, collocation, and networks of meaning interact to form a distinctive symbolic field. Second, from a methodological perspective, the integrated use of wordlists, concordances, and collocations proves to be an effective analytical tool for uncovering layers of meaning in drama scripts, so that it can be replicated in other texts that require simultaneous study of lexical structure and dramatic nuance. Third, in the practical realm, the results of this study can be utilized by scriptwriters, directors, and actors to manage word choice strategically—whether to build atmosphere, strengthen relationships between characters, or create deep emotional resonance for the audience.

BIBLIOGRAPHY

- Ahmad, J., & Herdiana, H. R. (2022). TINDAK TUTUR ILOKUSI DALAM NASKAH DRAMA MEGA-MEGA KARYA ARIFIN C. NOER. *Diksatrasia : Jurnal Ilmiah Pendidikan Bahasa Dan Sastra Indonesia*, 5(2). <https://doi.org/10.25157/diksatrasia.v5i2.7022>
- Ajmal, M., Rana, S., & Siddiqui, S. (2022). A Corpus-based Analysis of Lexical Cohesion in Charles Dickens's Great Expectations. *Global Language Review*, VII(III), 27–34. [https://doi.org/10.31703/glr.2022\(VII-III\).04](https://doi.org/10.31703/glr.2022(VII-III).04)
- Dmytruk, L. A., & Ostapenko, S. A. (2022). DRAMATICS TRANSLATION: LEXICO-SEMANTIC ASPECT. *INTELLIGENCE. PERSONALITY. CIVILIZATION*, 2 (25), 31–40. <https://doi.org/10.33274/2079-4835-2022-25-2-31-40>
- Fikri, H., & Hasanuddin WS, H. W. (2023). Pertentangan Kelas Sosial dalam Naskah Drama Kapai-Kapai Karya Arifin C. Noer. *Persona: Kajian Bahasa Dan Sastra*, 2(1), 75–85. <https://doi.org/10.24036/jpers.v2i1.124>
- Madieva, G., Orazbekova, I., & Tayeva, R. (2015). Lexical approach as a combination of applied linguistics, corpus linguistics and language teaching methodology. *INTED2015 Proceedings*, 6413-6420, 2015.
- Mukhlis, I. R. (2022). Literatur Review Pada Teknik Pendeteksi Ambiguitas Leksikal dalam Software Requirements Spesification. *Journal of Computer Science and Visual Communication Design*, 7(1), 66–79. <https://doi.org/10.55732/jikdiskomvis.v7i1.473>
- Nogier, J.-F., & Zock, M. (1992). Lexical choice as pattern matching. *Knowledge-Based Systems*, 5(3), 200–212. [https://doi.org/10.1016/0950-7051\(92\)90032-B](https://doi.org/10.1016/0950-7051(92)90032-B)
- Qin, W. (2023). Computer-assisted Lexical Analysis of ESL Writing Based on Cognitive Model. *Frontiers in Educational Research*, 6(1). <https://doi.org/10.25236/fer.2023.060113>
- Qiu, L. (2020). The Manner Result Complementarity and the Lexicalization Pattern of Chinese Motion Verbs. *Theory and Practice in Language Studies*, 10(12), 1598–1605. <https://doi.org/10.17507/tpls.1012.13>
- Safira, P., Khairunnisa, N., & Azura, S. (2024). Tokoh dan Penokohan pada Naskah Drama Mega-Mega Karya Arifin C. Noer: Analisis Hegemoni Gramsci. *Disastra: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 5(2), 222. <https://doi.org/10.29300/disastra.v5i2.3247>
- Safitri, R. N., Burhanuddin, B., & Saharudin, S. (2023). Metafora Leksikal dalam Teks

- Berita pada Situs Goal.com. *Kode: Jurnal Bahasa*, 12(3).
<https://doi.org/10.24114/kjb.v12i3.48278>
- Sánchez, D., Zunino, L., De Gregorio, J., Toral, R., & Mirasso, C. (2023). Ordinal analysis of lexical patterns. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(3). <https://doi.org/10.1063/5.0139852>
- Sidorova, E. A., Ivanov, A. I., & Ovchinnikova, K. A. (2025). Information extraction from texts based on ontology and large language models. *Ontology of Designing*, 15(1), 114–129. <https://doi.org/10.18287/2223-9537-2025-15-1-114-129>
- Turahmat, T. (2022). Nilai Religius dalam Naskah Drama “Sumur Tanpa Dasar” Karya Arifin C. Noer (Religious Values in Sumur Tanpa Dasar’s Drama Scripts By Arifin C. Noer). *Indonesian Language Education and Literature*, 7(2), 370. <https://doi.org/10.24235/ileal.v7i2.10393>
- Wu, S. (2024). An Analysis of Lexical Cohesion in the Text “After Twenty Years.” *Studies in Linguistics and Literature*, 8(3), p251. <https://doi.org/10.22158/sll.v8n3p251>