JPI | JURNAL PUSTAKA INDONESIA

OPEN ACCESS   Research Article

A Corpus-Informed Grammatical Profile of Indonesian Pop Lyrics: Word-Class Distribution and Denotation in Mahalini's *Fabula*

Nabila Nurrahma[1], Ali Akbar Jono[2,] and Vebi Andra[3]

[1,2,3] Universitas Islam Negeri Fatmawati Sukarno Bengkulu, Indonesia

ABSTRACT

Despite the widespread pedagogical use of songs, corpus-informed grammatical profiles of Indonesian pop lyrics remain scarce. This study therefore maps the distribution of Indonesian word classes and inventories denotative meanings in the official lyrics of the ten tracks on Mahalini's *Fabula*. Using a qualitative content-analysis design, we tokenized the corpus by orthographic word and annotated each token into one of ten classes (verba, adjektiva, nomina, pronomina, numeralia, adverbia, interogativa, demonstrativa, preposisi, konjungsi); denotative senses were then assigned with reference to KBBI VI Daring and resolved by local context. Descriptive statistics summarize the resulting profile of 323 lexical tokens and 157 denotative senses. Verbs dominate the lexicon (26.93%), followed by adjectives (15.79%) and pronouns (15.17%), with conjunctions (10.22%) and adverbs (8.98%) forming a mid-frequency band and nouns (6.19%), interrogatives (5.26%), demonstratives (4.64%), prepositions (4.02%), and numerals (2.79%) comparatively infrequent. The pattern indicates an affect-centred narrative style driven by processes, evaluative lexis, and deictic anchoring. The study contributes a replicable annotation protocol and a compact, clean dataset that can inform stylistic comparisons and instructional design. While limited to a single album and without inter-rater reliability or figurative/pragmatic layers, the findings provide a baseline for broader lyric corpora. Results guide Indonesian-language pedagogy toward high-yield resources (process verbs, evaluative adjectives, pronominal systems), support corpus-based stylistics across artists and subgenres, and offer benchmark material for Indonesian NLP in creative registers.

PUBLISHER'S NOTE

CORRESPONDING AUTHOR

Nabila Nurrahma, Universitas Islam Negeri Fatmawati Sukarno Bengkulu, Indonesia . Email: nabilanurrahma21@gmail.com

Introduction

Popular song lyrics constitute a hybrid register that fuses the patterned density of poetic language meter, rhyme, parallelism, and figurative tropes with the immediacy of colloquial discourse vernacular lexis, conversational syntax, deixis, and stance markers thereby offering a sensitive lens on how contemporary Indonesian is mobilized to construe experience, negotiate interpersonal relations, and index social positioning. Because lyrics circulate simultaneously as written text and performed speech, they foreground choices that cut across levels of grammar (from morphology and word class to clause structure and discourse organization) while

remaining anchored in everyday idiom. This dual character makes them an ideal testing ground for corpus-informed analysis of how meanings are built through recurrent grammatical resources, such as process verbs, evaluative adjectives, and pronominal address. Yet, despite a recent uptick in stylistic and corpus-based work that treats lyrics as a distinct text type, systematic grammatical profiling using transparent annotation schemes and comparable metrics across artists, subgenres, and periods remains comparatively sparse, leaving open the need for replicable descriptions that link lexical and morphosyntactic patterns to the communicative work lyrics routinely perform (Parada-Cabaleiro et al., 2024; Statham, 2020; Werner, 2021). Within Indonesian and broader applied-linguistics scholarship, songs are also widely recognised as authentic learning materials that heighten engagement and make linguistic structures salient, a conclusion reinforced by recent empirical and review studies showing positive effects of music-mediated learning on vocabulary, motivation, and broader academic outcomes (Anggaira et al., 2022; Chen et al., 2024; Vosniadou, 2013). Against this backdrop, mapping the lexical "building blocks" of lyrics specifically, the distribution of word classes and their literal (denotative) senses has dual value: it extends corpus-informed stylistics and provides a principled resource for curriculum designers who regularly adapt popular songs for donesian-language and literacy instruction (Parada-Cabaleiro et al., 2024; Statham, 2020; Werner, 2021).

From a linguistic standpoint, two considerations motivate our focus. First, cross-genre analyses consistently show that pop lyrics are characterized by unusually high rates of process encoding and evaluative coloring operationalized, respectively, as dense deployments of verbs and adjectives paired with equally dense deictic anchoring via first- and second-person pronouns. Verbs propel the storyline through sequences of actions, perceptions, and mental states (e.g., material, mental, and relational processes in a transitivity perspective), while adjectives supply the attitudinal load that nuances those processes with affective and axiological judgments. In parallel, pronominal choices ('I/you/we') and other deictic cues fix the interactional coordinates of the text, recruiting listeners into a here-and-now of address that feels intimate and direct. Taken together, these features foreground emotional trajectories shifts from longing to reconciliation, from doubt to commitment and sustain interpersonal address as the dominant discourse relation, a configuration that is especially typical of narrative love songs in which the chorus and verse structure repeatedly realigns stance, intensifies evaluation, and re-centers the dialogic bond between singer-persona and addressee (Parada-Cabaleiro et al., 2024; Statham, 2020; Werner, 2021). Second, at the semantic level, disaggregating denotation from connotation is necessary to avoid conflating conventional sense with evaluative overlays: denotation indexes the contextually appropriate dictionary-based meaning, whereas connotation covers culturally saturated associations and affective coloration (Brasoveanu & Rett, 2018; Castroviejo et al., 2021; Silk, 2021). Operationalising these distinctions within a transparent coding scheme aligns naturally with qualitative content analysis, which emphasises explicit category systems, systematic sampling/coding, and reliability-seeking procedures to enhance analytic trustworthiness (Bross, 2021; Hogeweg & Vicente, 2020; Wijaya & Zein, 2020).

In the Indonesian context, recent classroom-facing studies continue to demonstrate the pedagogical payoff of songs, including improved lexical uptake, sustained attention, and greater willingness to participate effects that further justify descriptive work capable of informing

materials development (Dewi Max et al., 2023; Niitemaa & Pietilä, 2018; Nkhobo & Chaka, 2021). At the same time, corpus and register-based approaches to lyrics have matured methodologically, offering replicable protocols for quantifying grammatical resources and comparing stylistic baselines across artists, subgenres, and eras (D'Ambrosio & Hedden, 2024; Silk, 2021; Wellwood, 2020). Taken together, these developments motivate a focused analysis of the grammatical composition of a coherent album, where token-level annotation of parts of speech can be combined with denotative glossing to produce a reusable descriptive profile.

While emerging Indonesian studies on Mahalini Raharja's *Fabula* have primarily examined pragmatic functions and semantic relations, there is so far as we can determine no corpus-informed, category by category account of the album's word-class distribution linked systematically to denotative meanings. Addressing this gap, the present study (i) describes the distribution of ten Indonesian word classes (verba, adjektiva, nomina, pronomina, numeralia, adverbia, interogativa, demonstrativa, preposisi, konjungsi) across *Fabula*'s official lyrics using a content-analytic protocol, and (ii) inventories the denotative senses instantiated in those categories to provide a replicable baseline for subsequent stylistic, pragmatic, and pedagogic inquiry (D'Ambrosio & Hedden, 2024; Goddard et al., 2019; Walkington et al., 2020).

Methods

This study employed a qualitative content-analysis design using the official lyrics of the ten tracks on Mahalini Raharja's *Fabula* (2023) as a bounded corpus. The dataset comprised only the album's released, full-length tracks; paratexts, metadata, and non-lexical vocalizations were excluded. Lyrics were retrieved from authorized sources and checked against publicly available official releases to ensure textual fidelity; the corpus was then normalized for orthography (capitalization and punctuation) while preserving original lexical forms. Tokenization proceeded by orthographic word. Following Indonesian grammatical scholarship (Cheung & Feng, 2021; Varnum et al., 2021; Werner, 2019), a coding manual defined ten word-class categories verba, adjektiva, nomina, pronomina, numeralia, adverbia, interogativa, demonstrativa, preposisi, and konjungsi with decision rules for borderline or multifunctional items (e.g., deverbal adjectives, pronominal clitics). Enclitic possessives/objects (-ku, -mu, -nya) were annotated as pronominal tokens and flagged as enclitics; contractions and colloquial spellings were resolved to their standard dictionary entries. Denotative meanings were identified as dictionary-grounded literal senses selected in context using KBBI VI Daring as the primary reference, supported where relevant by usage notes from standard grammars; polysemous items were assigned a single sense per token based on immediate cotext. Coding was iterative: one track was first piloted to calibrate operational definitions and refine the manual, after which the full corpus was annotated; problematic cases were resolved through analytic memoing and consensus checks recorded in an audit trail. The analytic strategy consisted of frequency counts by word class and by denotative sense and their conversion to percentages of the total token set; descriptive statistics and tabulations were produced with reproducible scripts in Python (pandas). Trustworthiness was enhanced through an explicit category system, consistency checks across tracks, and a documented workflow; inter-rater reliability was not calculated and is acknowledged as a limitation. As the study analyzes publicly accessible textual materials without human participants, institutional ethics review was not required.

Results and Discussion

Corpus Overview

The bounded corpus consists of the official lyrics to the ten tracks on *Fabula*. After orthographic normalization (capitalization and punctuation only), we identified 323 lexical tokens assigned to one of ten Indonesian word-class categories. Non-lexical vocalizations, paratexts, and metadata were excluded by design.

Distribusi of Word Classes

A transparent coding protocol yielded the frequency profile in Table 1, with verbs, adjectives, and pronouns forming the lexical backbone of the album. Verbs constitute the largest share (n = 87; 26.93%), followed by adjectives (n = 51; 15.79%) and pronouns (n = 49; 15.17%). Conjunctions (n = 33; 10.22%) and adverbs (n = 29; 8.98%) make up the mid-frequency band, while nouns (n = 20; 6.19%), interrogatives (n = 17; 5.26%), demonstratives (n = 15; 4.64%), prepositions (n = 13; 4.02%), and numerals (n = 9; 2.79%) occur less often. This pattern indicates a discourse style driven by processes and evaluations, with dense deictic anchoring through pronominal reference and relatively limited quantification.

Table 1. Distribution of Indonesian word classes in Fabula (N = 323)

Word class	Count	Share (%)
Verba	87	26.93
Ajektiva	51	15.79
Pronomina	49	15.17
Konjungsi	33	10.22
Adverbia	29	8.98
Nomina	20	6.19
Interogativa	17	5.26
Demonstrativa	15	4.64
Preposisi	13	4.02
Numeralia	9	2.79

Denotative Maeaning

Across the corpus we catalogued 157 denotative senses, defined as contextually appropriate dictionary-grounded meanings. Representative tokens per class are summarised in Table 2. These instances illustrate how the album's lexicon foregrounds actions and evaluations (e.g., membawa 'to carry', memutih 'to turn white'; pilu, bahagia) while pronominal clitics (-nya) serve possessive and object functions typical of intimate address in lyric discourse. Although the present study is descriptive, the distribution suggests that affective narration is supported by high-salience process lexemes and deictic framing, with comparatively fewer items devoted to quantification or complex spatial-temporal scaffolding (numerals, prepositions).

Table 2. Illustrative denotative senses by word class (examples)

Word class	Examples with denotative gloss
Verba	*membawa* 'to carry/transport'; *memutih* 'to turn white'
Ajektiva	*pilu* 'woeful'; *rindu* 'longing'; *bahagia* 'happy'
Nomina	*rambut* 'hair'; *hati* 'heart (seat of feelings)'

Word class	Examples with denotative gloss
Pronomina	*-nya* '3rd-person possessive/object enclitic'; *aku* 'I'; *kau/kamu* 'you'
Numeralia	*satu* 'one'
Adverbia	*pagi* 'morning'; *nanti* 'later'
Interogativa	*bagaimana* 'how'; *mengapa* 'why'
Demonstrativa	*itu* 'that'; *ini* 'this'
Preposisi	*dari* 'from'; *ke* 'to'
Konjungsi	*dan* 'and'; *atau* 'or'

Summary

Taken together, the counts and examples indicate a lexical profile centred on processes, evaluations, and deictic stance. The long-tail distribution of lower-frequency classes (numerals, prepositions, demonstratives) is consistent with a narrative-lyric genre that relies less on elaborated quantification and more on interpersonal positioning. These descriptive outcomes provide a replicable baseline for follow-up analyses (e.g., figurative language and speech-act functions) using the same coding protocol.

Discussion

The lexical profile of *Fabula* dominated by verbs (26.93%), adjectives (15.79%), and pronouns (15.17%) while numerals and prepositions remain comparatively sparse corroborates cross-register observations that mainstream pop lyrics systematically privilege processes, evaluations, and deictic stance over elaborated quantification or spatial scaffolding. In short, meanings are propelled by what happens, how it is appraised, and who is being addressed, rather than by fine-grained measurements or locative framing.

This distribution converges with Werner's register analysis, which documents high conversational density and sustained pronominal address as hallmarks of pop texts (Werner, 2021), and it aligns with Parada-Cabaleiro et al.'s cross-linguistic evidence that lyrics routinely foreground interpersonal proximity through first- and second-person deixis. It also echoes functional accounts of transitivity in lyric narratives wherein process types and evaluative lexis shoulder the thematic burden, giving narrative momentum to affect-centered storytelling.

Our denotative inventory extends this picture by supplying token-level, dictionary-grounded senses, thereby responding to stylistic syntheses that call for granular and replicable descriptions of lyric language beyond metaphor tallies or isolated figures of speech (Statham, 2021). At the same time, the profile bridges to education research on song-based instruction by offering an explicit map of "what words do" in authentic texts an asset for targeting lexical uptake, motivation, and task design (Biber & Egbert, 2018; Climent & Coll-Florit, 2021; Kadir, 2021; Nurhamidah et al., 2021).

Unlike prior Indonesian studies on *Fabula* that center pragmatic functions or relational semantics, the present work contributes a corpus-informed, category-by-category account coupling part-of-speech distribution with a transparent denotative coding protocol that can be reused for inter-album or cross-genre comparisons. Practically, the resulting lexical blueprint helps educators prioritize high-yield grammatical resources (e.g., process verbs, evaluative adjectives, pronominal systems) when crafting lyric-based tasks, gives songwriters and cultural analysts a measurable baseline for affect-oriented narrative style, and provides NLP/corpus

researchers with a small yet cleanly annotated dataset for benchmarking Indonesian POS tagging or sense selection in creative registers.

We acknowledge, however, that the corpus is restricted to a single artist and album; inter-rater reliability was not computed; orthographic tokenization under-represents multiword expressions; and the focus on denotation sidelines connotation, figurative force, and discourse-pragmatic functions, with no inferential modeling or cross-genre baselines included. Future work should broaden sampling across artists and subgenres, institute dual-coder reliability checks, integrate figurative/pragmatic layers, and test whether the observed distributional profile is stable in larger corpora and responsive to listening-dependent variables such as tempo, prosody, and performance context.

Conclusion

This study provides a corpus-informed grammatical profile of Mahalini's *Fabula* that links part-of-speech distributions to denotative meanings within a transparent content-analytic protocol. Annotating 323 lexical tokens across ten Indonesian word classes revealed a discourse style anchored in processes, evaluations, and deictic stance evidenced by the predominance of verbs (26.93%), adjectives (15.79%), and pronouns (15.17%) with numerals and prepositions comparatively rare; an inventory of 157 denotative senses further clarifies the literal semantic resources that scaffold the album's affect-centred narratives. Beyond documenting a recent pop corpus, the study contributes a replicable workflow and a small, clean dataset that can support pedagogical design (targeting high-yield grammatical resources in lyric-based tasks), stylistic comparison across artists and subgenres, and benchmarking for Indonesian NLP in creative registers. While the analysis is limited to a single album, relies on orthographic tokenization, and does not include inter-rater reliability or figurative/pragmatic layers, the protocol and findings offer a productive baseline. Future work should broaden sampling, compute reliability metrics, incorporate multiword expressions and prosodic/performative cues, and model cross-genre contrasts to test the stability and explanatory power of the observed lexical profile.

References

Anggaira, A. S., Aryanti, N., Suryadi, S., & Tusriyanto, T. (2022). Songs for teaching vocabulary: English learning media for preschoolers. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini, 6*(6), 6069–6078. https://doi.org/10.31004/obsesi.v6i6.2870

Biber, D., & Egbert, J. (2018). *Register variation online*. Cambridge University Press. https://doi.org/10.1017/9781316388228

Brasoveanu, A., & Rett, J. (2018). Evaluativity across adjective and construction types: An experimental study. *Journal of Linguistics, 54*(2), 263–329. https://doi.org/10.1017/S0022226717000123

Bross, F. (2021). On the interpretation of expressive adjectives: Pragmatics or syntax? *Glossa: A Journal of General Linguistics, 6*(1). https://doi.org/10.5334/gjgl.1214

Castroviejo, E., Fraser, K., & Vicente, A. (2021). More on pejorative language: Insults that go beyond their extension. *Synthese, 198*(10), 9139–9164. https://doi.org/10.1007/s11229-020-02624-0

Chen, M., Mohammadi, M., & Izadpanah, S. (2024). Language learning through music on the academic achievement, creative thinking, and self-esteem of the English as a foreign language (EFL) learners. *Acta Psychologica, 247*, 104318. https://doi.org/10.1016/j.actpsy.2024.104318

Cheung, J. O., & Feng, D. (William). (2021). Attitudinal meaning and social struggle in heavy metal song lyrics: A corpus-based analysis. *Social Semiotics, 31*(2), 230–247. https://doi.org/10.1080/10350330.2019.1601337

Climent, S., & Coll-Florit, M. (2021). All you need is love: Metaphors of love in 1946–2016 Billboard year-end number-one songs. *Text & Talk, 41*(4), 469–491. https://doi.org/10.1515/text-2019-0209

D'Ambrosio, J., & Hedden, B. (2024). Multidimensional adjectives. *Australasian Journal of Philosophy, 102*(2), 253–277. https://doi.org/10.1080/00048402.2023.2277923

Dewi Max, J. I. S., Vivian, Y. I., Putra, B. A., & Kuncara, S. D. (2023). Transitivity analysis of Taylor Swift's lyrics in *Midnights* album. *Indonesian Journal of EFL and Linguistics*, 279–294. https://doi.org/10.21462/ijefl.v8i2.688

Goddard, C., Taboada, M., & Trnavac, R. (2019). The semantics of evaluational adjectives. *Functions of Language, 26*(3), 308–342. https://doi.org/10.1075/fol.00029.god

Hogeweg, L., & Vicente, A. (2020). On the nature of the lexicon: The status of rich lexical meanings. *Journal of Linguistics, 56*(4), 865–891. https://doi.org/10.1017/S0022226720000316

Kadir, R. (2021). Code-switching in Indonesian popular songs and the implications for English language teaching. *JOALL (Journal of Applied Linguistics and Literature), 6*(1), 109–132. https://doi.org/10.33369/joall.v6i1.13314

Niitemaa, M.-L., & Pietilä, P. (2018). Vocabulary skills and online dictionaries: A study on EFL learners' receptive vocabulary knowledge and success in searching electronic sources for information. *Journal of Language Teaching and Research, 9*(3), 453. https://doi.org/10.17507/jltr.0903.02

Nkhobo, T., & Chaka, C. (2021). Exploring instances of Deleuzian rhizomatic patterns in students' writing and in online student interactions. *International Journal of Learning, Teaching and Educational Research, 20*(10), 1–22. https://doi.org/10.26803/ijlter.20.10.1

Nurhamidah, I., Purwanto, S., & Wulandari, D. F. (2021). Alan Walker's "Lily" song lyrics to depict satanic tricks: A generic structure analysis as rhetoric. *JOALL (Journal of Applied Linguistics and Literature), 6*(2), 246–263. https://doi.org/10.33369/joall.v6i2.14204

Parada-Cabaleiro, E., Mayerl, M., Brandl, S., Skowron, M., Schedl, M., Lex, E., & Zangerle, E. (2024). Song lyrics have become simpler and more repetitive over the last five decades. *Scientific Reports, 14*(1), 5531. https://doi.org/10.1038/s41598-024-55742-x

Silk, A. (2021). Evaluational adjectives. *Philosophy and Phenomenological Research, 102*(1), 127–161. https://doi.org/10.1111/phpr.12635

Statham, S. (2020). The year's work in stylistics 2019. *Language and Literature: International Journal of Stylistics, 29*(4), 454–479. https://doi.org/10.1177/0963947020970658

Varnum, M. E. W., Krems, J. A., Morris, C., Wormley, A., & Grossmann, I. (2021). Why are song lyrics becoming simpler? A time series analysis of lyrical complexity in six decades of American popular music. *PLOS ONE, 16*(1), e0244576. https://doi.org/10.1371/journal.pone.0244576

Vosniadou, S. (2013). *International handbook of research*.

Walkington, Z., Wigman, S. A., & Bowles, D. (2020). The impact of narratives and transportation on empathic responding. *Poetics, 80*, 101425. https://doi.org/10.1016/j.poetic.2019.101425

Wellwood, A. (2020). Interpreting degree semantics. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02972

Werner, V. (2019). Assessing hip-hop discourse: Linguistic realness and styling. *Text & Talk, 39*(5), 671–698. https://doi.org/10.1515/text-2019-2044

Werner, V. (2021). Catchy and conversational? A register analysis of pop lyrics. *Corpora, 16*(2), 237–270. https://doi.org/10.3366/cor.2021.0219

Wijaya, V., & Zein, T. T. (2020). Visual and verbal in Line Messenger application: A semiotic study. *JOALL (Journal of Applied Linguistics & Literature), 5*(2), 213–222. https://doi.org/10.33369/joall.v5i2.10268