IKR Publishers

IKR Journal of Education and Literature (IKRJEL)
Journal homepage: https://ikrpublishers.com/ikrjel/
Volume-2, Issue-6 (November-December) 2026

OPEN ACCESS
ISSN: 3107-7684 (Online)

A Morphological Analysis of Compound Words in Mambwe: Types, Prevalence and Productivity in a Zambian Bantu Language

Norah Nampasa[1], Pethias Siame[2*] & Kennedy N. Kangwa[3]
[1,2,3]Department of Literature and Languages, Kwame Nkrumah University, Zambia

DOI: *10.5281/zenodo.24446721*

ARTICLE INFO

Article history:
Received : 25-10-2026
Accepted : 07-11-2026
Available online : 22-11-2026

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial **4.0 International License (CC BY-NC 4.0),** which permits non-commercial use, sharing, and reproduction in any medium, provided the original author(s) and source are appropriately credited.

Citation: Nampasa, N., Siame, P., & Kangwa, K. N. (2026). A Morphological Analysis of Compound Words in Mambwe: Types, Prevalence and Productivity in a Zambian Bantu Language. *IKR Journal of Education and Literature (IKRJEL), 2*(6), 1-9.

ABSTRACT **Original Research Article**

This study identifies the types of compound words most prevalent in the Mambwe language of Mbala District, Zambia, a Bantu language classified as zone M.15 whose compounding system had not previously been documented. Grounded in Lexical Morphology Theory, and in particular its principles of lexical levels and productivity, the research adopted a qualitative approach within an interpretivist paradigm and an exploratory design. Ten native speakers aged between 30 and 65 years were selected through purposive and snowball sampling, and data were collected through semi-structured interviews, a compound elicitation matrix organised by semantic domain, focus group discussions, and participant observation. The data were morphologically segmented and analysed thematically, yielding a corpus of twenty-eight compound forms distributed across five structural schemas. Results revealed that Noun + Noun compounds are the most prevalent type, accounting for 35.7 percent of the corpus, followed by Noun + Adjective and Noun + Verb compounds at 21.4 percent each, while multi-constituent associative and reduplicative compounds account for 10.7 percent each. The three nominal schemas together account for 67.9 percent of the corpus, confirming that Mambwe compounding is predominantly a nominal operation. Eight of the ten Noun + Noun compounds are endocentric, showing that the most frequent type is also the most semantically transparent, the profile Lexical Morphology Theory predicts for productive Level II word formation. The study concludes that compounding remains a productive source of new vocabulary in Mambwe, and recommends the documentation of its less frequent and culturally specialised compound types, which younger speakers know least well.

Keywords: Compound Words, Mambwe Language, Lexical Morphology Theory, Morphological Productivity, Bantu Morphology, Zambia.

**Corresponding author: Pethias Siame*
Department of Literature and Languages, Kwame Nkrumah University, Zambia

Introduction

Compounding, the process by which two or more independent words combine to form a single lexical unit, is one of the principal means by which Bantu languages expand their vocabulary to meet new communicative needs. The Mambwe language, classified under Guthrie's (1948) zone M15, is spoken in Mbala and Senga districts of Northern Province, Zambia, and in parts of south-western Tanzania (Siame, 2023). As a member of the Niger-Congo family, Mambwe exhibits the characteristic Bantu features of an extensive noun class system and agglutinative morphology, features that shape how words combine to form compounds. The language carries the cultural heritage and traditional knowledge of the Mambwe people, much of which is encoded in compound vocabulary that names kinship relations, social roles, moral qualities, and ritual concepts.

Morphological research on Bantu languages has traditionally concentrated on the verb, whose agglutinative structure lends itself readily to formal analysis, while nominal compounding has received comparatively little attention (Nurse & Philippson, 2003). Research on Mambwe has followed the same pattern: existing studies have examined tense and verbal extensions (Siame, 2023; Siame et al., 2023; Siame & Banda, 2024), but no study has established which types of compound

words occur in the language, or which of them are most prevalent. This omission matters for two reasons. Theoretically, the relative frequency of compound types indicates which word-formation schemas remain productive in a language. Practically, compound vocabulary that is not documented is vulnerable to loss as language shift towards English and Bemba reduces the domains in which Mambwe is used.

This study presents the morphology of compound words in Mambwe by addressing the broader objective, namely to identify the types of compound words most prevalent in the Mambwe language. It addresses the research question: What types of compound words are most prevalent in the Mambwe language? By establishing the structural types of Mambwe compounds, measuring their relative frequency, and interpreting the resulting distribution through Lexical Morphology Theory, the study provides an empirical baseline for the documentation of Mambwe. It contributes data from an under-described language to the comparative study of Bantu compounding.

The significance of this study extends beyond descriptive documentation (Siame & Lubbungu, 2023). As Batibo (2005) observes, the specialised vocabulary of African languages is among the first to be lost under conditions of language shift, and compounding represents a particularly vulnerable domain because compound meanings often depend on cultural knowledge that is not transparent to younger speakers. By identifying which compound types are most prevalent and which are most at risk, this study offers practical guidance for language documentation efforts and for the development of pedagogical materials that treat productive compound schemas as patterns for lexical innovation rather than as fixed inventories (Siame, 2022; Siame et al., 2025).

The article is organised as follows. Section 2 reviews the relevant literature on compound typology and Bantu compounding. Section 3 presents the theoretical framework. Section 4 describes the methodology, including the research design, sampling, data collection, and analysis procedures. Section 5 presents the results and discusses them in relation to the literature. Section 6 concludes with recommendations for documentation, pedagogy, and future research.

Literature Review

Global Frameworks for Compound Classification

Global scholarship has developed systematic frameworks for identifying and classifying compound words. Scalise and Vogel (2010) present a cross-linguistic typology of compounds that distinguishes compound types according to the grammatical relation between their constituents, while Guevara and Scalise (2009) set out criteria for distinguishing compounds from derived words and phrasal combinations that superficially resemble them. Lieber (2009) advances methods for categorising compounds according to the semantic relationship between their constituents, and Hay and Baayen (2005) demonstrate that the prevalence of a morphological type correlates with its productivity, so that the most frequent compound types in a language are typically those that speakers continue to use to create new words. These frameworks provide valuable tools for identifying prevalent compound types, but they were developed largely on Indo-European data and do not anticipate how a language organised around obligatory noun class concord distributes its compound types.

The distinction between endocentric and exocentric compounds, first articulated by Bloomfield (1933), remains fundamental to compound classification. Endocentric compounds have a semantic head that determines both the category and the core meaning of the whole, as in English *blackbird*, where *bird* identifies what kind of creature is denoted. Exocentric compounds, by contrast, lack a semantic head within the compound itself, so that their meaning cannot be derived from the constituents alone, as in English *redhead*, which denotes a person rather than a type of head. This distinction is particularly relevant to Bantu languages, where noun class agreement provides formal evidence of headedness that may either align with or diverge from semantic headedness.

Bantu Compounding

African scholarship provides more direct precedents for the study of Bantu compounding. Katamba (2003), drawing extensively on Luganda, identifies Noun + Noun compounding as the most productive compound type, attributing this to the rich foundation that the noun class system provides for nominal concatenation, and argues that Bantu compounds require language-specific categorisation rather than categories imported from European languages. Nurse and Philippson (2003) describe nominal compounding as a primary mechanism of lexical expansion across the Bantu family, generating terms for kinship, social roles, and material culture, while noting that it has been studied far less than verbal morphology. Basciano et al. (2011), in a comparative study of Bemba, Italian, and Mandarin Chinese, establish that Bantu nominal compounds are characteristically left-headed and formed by the concatenation of nominal stems, but conclude that Noun + Noun compounding in Bemba is largely unproductive.

Because Bemba is geographically and genetically close to Mambwe, this conclusion raises the question of whether the same holds for Mambwe. Batibo (2005) adds that such questions are urgent, since the specialised vocabulary of African languages is among the first to be lost under language shift. The Bantu noun class system provides both opportunities and constraints for compounding: noun class prefixes on compound constituents must be compatible, and agreement patterns within compounds reveal the headedness relations that operate in the language. As Katamba (2003)

notes, the noun class system creates a rich morphological foundation for nominal compounding, but it also imposes structural requirements that distinguish Bantu compounds from their Indo-European counterparts.

Zambian Research on Mambwe and Neighbouring Languages

Zambian research has begun to document the morphology of Mambwe and its neighbours. Siame (2023) provides a morphological analysis of tense in Mambwe, and Siame and Banda (2024) analyse micro-variations in verbal extensions across ciLungu, ciMambwe, and ciNamwanga, establishing the noun class and agreement patterns against which compound structure can be described. Beyond Mambwe, Zimba (2014) examines the grammar of the compound noun in Tumbuka, demonstrating that Zambian Bantu compounding rewards systematic investigation. However, the existing work on Mambwe is oriented almost entirely towards the verb, and no study has identified the types of compound words that occur in the language or measured their relative prevalence, representing a significant gap this study addresses.

Summary of Literature and Research Gap

The literature reviewed above establishes three points that inform the present study. First, while global frameworks for compound classification provide useful analytical categories, they require adaptation to the morphological realities of Bantu languages, particularly the noun class system and its implications for headedness and agreement. Second, existing research on Bantu compounding, particularly Katamba's (2003) work on Luganda and Basciano et al.'s (2011) study of Bemba, establishes that Noun + Noun compounding is a significant word-formation strategy in the family, though its productivity may vary across languages. Third, research on Mambwe has focused overwhelmingly on verbal morphology, leaving nominal compounding undocumented and unanalysed.

This study addresses that gap by providing the first systematic description of compound types in Mambwe, measuring their relative frequency, classifying them according to headedness, and interpreting the results through Lexical Morphology Theory. In doing so, it contributes both to the descriptive documentation of an under-described language and to the comparative study of Bantu compounding.

Theoretical Framework

Lexical Morphology Theory

This research is grounded in Lexical Morphology Theory, proposed by Kiparsky (1982), which holds that word formation takes place in the lexicon through an ordered series of levels, each associated with its own morphological and phonological rules. The theory rests on several principles: lexical levels, the Lexical Integrity Hypothesis, affixation and rule ordering, the interaction of morphology with phonology, and productivity. Although Kiparsky's original formulation was concerned principally with derivation, the framework was subsequently extended to compounding by Booij (2012), who treats compounding as a Level II operation applying to already-formed words, and by Katamba (1993, 2003), who shows that the level distinction can be maintained in Bantu languages, where noun class prefixes on compound constituents behave as Level II material.

Productivity and Transparency

The principle most directly relevant to the present objective is productivity. On the extended version of the theory, Level II processes are loosely joined, regular, and semantically transparent, whereas forms whose meanings cannot be recovered from their constituents are stored in the lexicon as conventionalised items in the manner characteristic of Level I. The theory therefore predicts that the most productive compound schemas in a language will also be the most frequent and the most transparent, because speakers continue to use them to create new words whose meanings other speakers can readily interpret. The distinction between endocentric compounds, whose meaning is headed by one of their constituents, and exocentric compounds, whose meaning lies outside their constituents (Bloomfield, 1933), provides a measure of that transparency.

Application to the Present Study

Applied to the present study, Lexical Morphology Theory guided both the identification and the interpretation of prevalent compound types. Each compound was assigned to a structural schema and classified as endocentric or exocentric, and the frequency of each schema was then related to its transparency in order to establish whether the compound types that dominate Mambwe are those the theory predicts to be productive at Level II. Construction Grammar (Goldberg, 2006) was used in a supporting role to account for the conventionalised meanings of exocentric forms, which a purely rule-based account leaves unexplained.

The theoretical framework generates two predictions that the study tests. First, the most frequent compound schema should be the most transparent, exhibiting a high proportion of endocentric forms whose meanings are recoverable from their constituents. Second, the less frequent schemas should include a higher proportion of exocentric forms whose meanings depend on cultural or conventionalised knowledge. If the Mambwe data confirm these predictions, they provide support for the extension of Lexical Morphology Theory to Bantu compounding; if they do not, they raise questions about the universality of the Level I/Level II distinction in languages with noun class systems.

Methodology

Research Design and Paradigm

This study adopted a qualitative approach within an interpretivist paradigm, which was suited to understanding how Mambwe speakers construct compound words and what those constructions signify within their cultural context (Creswell & Creswell, 2023). An exploratory research design was employed because the compounding patterns of Mambwe had not previously been documented, and the study sought to describe them in their natural setting without manipulating any variables.

The interpretivist paradigm was appropriate for this research because compound words are not merely formal structures but carriers of cultural meaning. As Creswell and Creswell (2023) argue, qualitative research within an interpretivist paradigm seeks to understand phenomena from the perspective of participants, recognising that meaning is constructed through social interaction. Compound words in Mambwe encode cultural knowledge about kinship, social roles, and ritual concepts, and understanding their prevalence required attention to the meanings participants attributed to them, not merely to their formal structure.

Sampling and Participants

The target population comprised adult Mambwe speakers aged between 30 and 65 years residing in Mbala District. Participants were selected through purposive sampling, which targeted speakers known in their communities for linguistic competence and cultural knowledge, and through snowball sampling, in which participants recommended other knowledgeable speakers (Kumar, 2019; Okolie et al., 2025). The sample consisted of ten native speakers, five men and five women, including traditional elders, a village headman, a village counsellor, a traditional healer, a cultural custodian, teachers, a church leader, a farmer, and a trader. Five participants were rated as having high competence in the traditional register of ceremonial and culturally specialised vocabulary, and five as having moderate competence.

The sample size was determined by data saturation, defined as the point at which three consecutive elicitation sessions yielded no compound type not already recorded; saturation was reached at the eighth participant, and the remaining two sessions served to confirm the corpus. As Kumar (2019) notes, purposive sampling is appropriate when the researcher seeks participants with specific knowledge or characteristics relevant to the research question, while snowball sampling is useful for accessing populations that are difficult to identify through other means. The combination of both strategies ensured that participants included both community-recognised linguistic experts and ordinary speakers whose usage reflects everyday patterns.

Data Collection Instruments

Data were collected through four instruments, refined through a pilot study with three speakers. A semi-structured interview guide explored compound word usage in daily and cultural life, with each interview lasting between thirty and forty-five minutes. A compound elicitation matrix, organised by semantic domain to cover kinship, the natural environment, material culture, human attributes, and the ritual and spiritual sphere, prompted participants to produce compound words and to explain their parts and meanings. Two focus group discussions, one with participants of different ages and one with cultural experts, validated the elicited forms and their meanings, and participant observation at community gatherings documented compound usage in natural conversation. Sessions were audio-recorded in WAV format and supplemented by field notes.

The semi-structured interview format allowed participants to elaborate on compound meanings and usage contexts that a fixed questionnaire would have constrained. As Kumar (2019) observes, semi-structured interviews are particularly suited to exploratory research because they permit the researcher to pursue unexpected lines of inquiry while maintaining focus on the research objectives. The compound elicitation matrix served as a systematic prompt that ensured coverage of all relevant semantic domains, while the focus group discussions provided opportunities for participants to negotiate and confirm meanings collectively. Participant observation, conducted at community gatherings including weddings, funerals, and traditional ceremonies, allowed the researcher to document compound usage in natural contexts that might not have been elicited in formal interviews.

Data Analysis Procedures

Data analysis proceeded in three phases. In the first phase, recordings were transcribed and verified with a first-language research assistant, and an inventory of candidate compounds was assembled. In the second phase, each compound was segmented into its constituent morphemes, its constituents were assigned to word classes and noun classes, and the compound was assigned to a structural schema and classified as endocentric or exocentric. In the third phase, the segmented corpus was coded thematically, with the first theme addressing the prevalence of compound types; frequencies and percentages for each schema were computed in R. The research assistant independently coded the full corpus for structural type, and points of divergence were resolved by reference to the classificatory criteria rather than in favour of either coder. Preliminary analyses were returned to participants for member checking.

The morphological segmentation followed the conventions established in the Bantu literature, with noun class prefixes identified according to the standard numbering system and verbal extensions analysed according to the patterns described by Siame and Banda (2024). The classification of compounds as endocentric or exocentric followed Bloomfield's (1933) criteria, with the semantic head identified by asking whether the compound denotes a subtype of the category named by one of its constituents. Inter-coder reliability was assessed by comparing the research assistant's

independent classifications with those of the primary researcher, and any discrepancies were resolved through discussion and reference to the classificatory criteria.

Ethical Considerations

Ethical clearance was not obtained from the Nkrumah University Research Ethics Committee. Nevertheless, permission to conduct the research was secured from traditional leaders in Mbala District. Written informed consent was obtained before any recording, and participants were informed of their right to withdraw at any stage without consequence. Participants were identified only by codes (P01 to P10) to preserve confidentiality. Because some compounds belong to ritual and spiritual contexts, ritual vocabulary was not recorded without explicit permission, and participants retained the right to withhold any form without giving a reason.

The decision to seek permission from traditional leaders before approaching individual participants reflects the communal decision-making structures that govern research in rural Zambian communities. As Kumar (2019) notes, ethical research practice requires attention to local norms and values, not merely to formal institutional requirements. The provision allowing participants to withhold ritual vocabulary without giving a reason recognised the sensitivity of culturally specialised knowledge and the right of knowledge holders to control its dissemination.

Results and Discussion

This section presents findings on the types of compound words most prevalent in the Mambwe language of Mbala District. Compounds are cited in the practical orthography of Mambwe, with constituents separated by a plus sign and English glosses given in single quotation marks.

Distribution of Compounds

Analysis of the data yielded a corpus of twenty-eight compound forms distributed across five structural schemas: Noun + Noun, Noun + Adjective, Noun + Verb, multi-constituent associative, and reduplicative compounds. Table 1 presents the distribution of these forms by structural schema and headedness. Noun + Noun compounds form the single largest category, accounting for ten of the twenty-eight forms (35.7 percent), followed by Noun + Adjective and Noun + Verb compounds with six forms each (21.4 percent), and by multi-constituent and reduplicative compounds with three forms each (10.7 percent). The three nominal schemas, namely Noun + Noun, Noun + Adjective, and multi-constituent associative compounds, together account for nineteen of the twenty-eight forms (67.9 percent), indicating that Mambwe compounding is predominantly a nominal operation. Across the corpus, eighteen forms (64.3 percent) are endocentric, and ten (35.7 percent) are exocentric. Representative examples of each schema are presented in Table 2.

Table 1: Distribution of Mambwe Compounds by Structural Schema and Headedness (N = 28)

Structural Schema	Endocentric	Exocentric	Total	%
Noun + Noun	8	2	10	35.7
Noun + Adjective	4	2	6	21.4
Noun + Verb (relativised or derived)	3	3	6	21.4
Multi-constituent (associative)	0	3	3	10.7
Reduplicative	3	0	3	10.7
Total	**18**	**10**	**28**	**100.0**

It should be noted that in endocentric compounds, the meaning of the compound is headed by one of its constituents, while in exocentric compounds, the meaning of the compound lies outside its constituents. In addition, the percentages presented represent the full corpus.

The structural schema of compounds in the Mambwe language are shown below:

Table 2: Representative Mambwe Compounds by Structural Schema

Schema	Compound	Segmentation	Gloss	Headedness
Noun + Noun	*Mwine-muzi*	mwine + muzi	Village headman	Endocentric
Noun + Noun	*Taata-vyala*	taata + vyala	Father-in-law	Endocentric
Noun + Noun	*Ikanda-ngombe*	i-kanda + ngombe	Heavy rain	Exocentric
Noun + Adjective	*Muntu-musuma*	muntu + musuma	Good person	Endocentric
Noun + Adjective	*Kacema-musuma*	kacema + musuma	Good Shepherd	Exocentric
Noun + Verb	*Ifwa-linowalya*	ifwa + li-no-walya	Death that eats	Endocentric
Noun + Verb	*Yemba-wacilolwa*	yemba + wa-ci-lolwa	Sea of glass	Exocentric
Multi-constituent	*Mama-waluse*	mama + wa + luse	Merciful woman	Exocentric
Reduplicative	*Ukukoleka-koleka*	uku-koleka + koleka	Perpetually drunk	Endocentric

It is worth noting that in Table 2, a plus sign separates the constituents of a compound and a hyphen separates morphemes within a constituent.

The distribution presented in Table 1 provides the first empirical answer to the research question, establishing that Noun + Noun compounds are the most prevalent type in Mambwe, followed by Noun + Adjective and Noun + Verb compounds, with multi-constituent associative and reduplicative compounds being the least frequent. This ordering aligns with the pattern that Katamba (2003) documents for Luganda, where Noun + Noun compounding is identified as the most productive type, and with Nurse and Philippson's (2003) observation that nominal compounding is a primary mechanism of lexical expansion across the Bantu family. The predominance of nominal schemas, which together account for 67.9 percent of the corpus, confirms that Mambwe compounding is primarily a nominal operation, consistent with the typological profile of Bantu languages described in the literature.

Noun + Noun Compounds

Noun + Noun compounds emerged as the most prevalent type of compound word in Mambwe. They are formed through the direct concatenation of two nominal bases, with the semantic head typically in the first position, consistent with the head-initial pattern of Bantu nominal morphology. The schema is the most consistently transparent in the corpus: eight of its ten forms (80 percent) are endocentric. *Mwine-muzi* (*mwine* 'owner, lord' + *muzi* 'village') 'village headman' combines a Class 1 noun of proprietorship with a Class 3 noun denoting a settlement, and its head *mwine* determines that the compound refers to a person. The kinship pair *Taata-vyala* 'father-in-law' and *Nyoko-vyala* 'mother-in-law' combines a parental term with *vyala*, a nominalised form of the verb *-vyala* 'to give birth', and shows that the schema operates productively: substituting the parental term generates a new kinship term within the same pattern. The pair *Umwana-mwanaci* 'baby girl' and *Umwana-monsi* 'baby boy' similarly specifies the head *umwana* 'child' by gender, while *Mwene-mukulu* 'great chief' and *Icintu-wingi* 'crowd or group of people' use a second constituent to intensify the head.

The productivity of the Noun + Noun schema is evident in the ease with which substituting one constituent generates a new compound within the same pattern. This observation supports Katamba's (2003) argument that the noun class system provides a rich foundation for nominal concatenation in Bantu languages, and it aligns with Hay and Baayen's (2005) finding that the prevalence of a morphological type correlates with its productivity. The fact that Noun + Noun compounds were elicited from all participant groups, regardless of age or competence level, further indicates that this schema remains robustly productive across the speech community.

Only two Noun + Noun compounds are exocentric. *Ikanda-ngombe* (*ikanda* 'hide' + *ngombe* 'cattle') denotes heavy, torrential rain, a meaning that rests on the image of earth saturated like a soaked cattle hide and cannot be recovered from the constituents without cultural knowledge. *Kaa-kolwe* 'small agile monkey' is extended figuratively to a quick or agile person, shifting its referent outside the compound. The classification of *Kaa-kolwe* as Noun + Noun is provisional, since *kolwe* carries adjectival meaning but appears without adjectival agreement morphology; reclassification would alter the count for the schema by one form. Noun + Noun compounds were also the most widely shared type, being elicited from all participant groups, and they serve the broadest range of everyday referential functions, naming kinship relations, social roles, natural phenomena, and personal characteristics.

The exocentric examples illustrate the role of cultural knowledge in interpreting compound meanings. As Goldberg (2006) argues, conventionalised meanings that cannot be derived from constituents alone are stored as constructions, and the meaning of *Ikanda-ngombe* depends on cultural knowledge about cattle hides and rain that is not encoded in the compound's form. This observation supports the use of Construction Grammar as a supplementary framework for explaining exocentric compounds, as noted in the theoretical framework.

Noun + Adjective Compounds

Noun + Adjective compounds were the second most prevalent type, accounting for six forms (21.4 percent), four of them endocentric. In these compounds, the adjective follows the noun it modifies and agrees with it in noun class, as in *Muntu-musuma* (*muntu* 'person' + *musuma* 'good') 'good person', *Mwana-mwipe* 'bad child', and *Cimela-cinono* 'small growing thing, seedling', in which the Class 7 prefix ci- appears on both constituents. *Imbuzi-yalwango* 'hungry goat' links its head to the quality of hunger through the Class 9 associative marker ya-. The schema clusters around a small set of evaluative meanings, principally goodness, badness, and smallness, reflecting the small, closed class of true adjectives in Mambwe, a feature typical of Bantu languages (Katamba, 2003). Its two exocentric forms, *Musamaliya-musuma* 'Good Samaritan' and *Kacema-muzuma* 'Good Shepherd', are of particular interest because older participants produced them spontaneously to name referents introduced through religious contact, providing direct evidence that the schema is still used to create new words.

The observation that the Noun + Adjective schema clusters around a limited set of evaluative meanings is consistent with Katamba's (2003) description of the adjective class in Bantu languages as small and closed. The exocentric forms *Musamaliya-musuma* and *Kacema-musuma* are particularly significant because they demonstrate that the schema is not merely a relic of traditional vocabulary but remains available for lexical innovation. The fact that these compounds were coined to name referents introduced through religious contact shows that speakers draw on productive schemas to

accommodate new concepts, a pattern that Hay and Baayen (2005) identify as characteristic of productive morphological processes.

Noun + Verb Compounds

Noun + Verb compounds also accounted for six forms (21.4 percent), but unlike the preceding schemas, they are evenly divided between three endocentric and three exocentric forms. These compounds typically combine a nominal head with a relativised or nominalised verb. In *Ifwa-linowalya* (*ifwa* 'death' + *li-no-walya* 'which eats'), 'death that eats', the Class 5 relative prefix li- agrees with the Class 5 head *ifwa*, showing that agreement is maintained across the compound boundary. *Yemba-wacilolwa* 'sea of glass', built on the passive relative form of *-lola* 'to see', and *Indya-muluwa* 'spiritual nourishment' are exocentric, their meanings depending on cosmological and ritual knowledge. The even split between transparent and opaque forms suggests that compounds containing a verbal constituent are more likely than nominal compounds to acquire meanings that go beyond their parts.

The maintenance of noun class agreement across the compound boundary in *Ifwa-linowalya* provides formal evidence that the compound is a single lexical unit rather than a phrase, supporting the classification of such forms as compounds rather than syntactic constructions. This observation aligns with Guevara and Scalise's (2009) criteria for distinguishing compounds from phrases, which include the presence of morphological integration between constituents. The higher proportion of exocentric forms in the Noun + Verb schema, compared with the Noun + Noun schema, suggests that verbal constituents contribute to semantic opacity, perhaps because the relativised verb forms encode complex propositions whose interpretation depends on context and cultural knowledge.

Multi-constituent and Reduplicative Compounds

Multi-constituent associative and reduplicative compounds were the least prevalent types, each accounting for three forms (10.7 percent), and they occupy opposite ends of the transparency scale. All three multi-constituent compounds are exocentric. In these forms, two nominal bases are joined by the associative marker wa-, as in *Mupashi-wamuzilo* (*mupashi* 'spirit, breath' + wa- + *muzilo* 'holiness') 'Holy Spirit', *Mupasi-wamuzilo* 'spirit of sacred breath', and *Mama-waluse* (*mama* 'mother' + wa- + *luse* 'mercy') 'merciful woman'. These compounds encode dense theological and evaluative meanings and were produced principally by participants with high traditional register competence. By contrast, all three reduplicative compounds are endocentric. *Ukutunda-tunda* 'frequent urination', *Ukukoleka-koleka* 'being perpetually drunk', and *Zuwa-zuwa* 'very fast' repeat a verbal or adverbial base to express repeated, habitual, or intensified action, and their meanings remain recoverable from the repeated base.

The concentration of exocentricity in the multi-constituent associative schema, and its association with participants of high traditional register competence, supports the theoretical prediction that less frequent schemas are more likely to involve conventionalised meanings that depend on cultural knowledge. As Goldberg (2006) argues, such conventionalised form-meaning pairings are stored as constructions in the speaker's lexicon, and their acquisition requires exposure to the cultural contexts in which they are used. The reduplicative schema, by contrast, remains semantically transparent despite its low frequency, suggesting that frequency and transparency are not perfectly correlated across all schema types. This observation qualifies the predictions of Lexical Morphology Theory and suggests that reduplication may constitute a distinct type of word-formation process with its own transparency profile.

Productivity, Transparency and Generational Distribution

Taken together, the findings support the productivity principle of Lexical Morphology Theory. The theory predicts that productive compounding operates at Level II, where constituents are loosely joined, and the resulting words are regular and transparent (Booij, 2012; Kiparsky, 1982). In Mambwe, the most frequent schema, Noun + Noun, is also the most consistently transparent, with eight of its ten forms endocentric, whereas exocentricity is concentrated in the less frequent associative and verbal schemas whose meanings depend on cultural knowledge. This convergence of frequency and transparency is consistent with Hay and Baayen's (2005) finding that the prevalence of a morphological type reflects its productivity. The ordering of types observed here, in which Noun + Noun outranks Noun + Adjective and Noun + Verb, which in turn outrank multi-constituent and reduplicative forms, parallels the pattern Katamba (2003) documents for Luganda, suggesting a tendency shared across the Bantu family. It contrasts, however, with the conclusion of Basciano et al. (2011) that Noun + Noun compounding is largely unproductive in Bemba; whether this reflects a genuine difference between the two languages or a difference in the registers sampled is a question the present data cannot settle.

The distribution of compound types across speakers points in the same direction. Participants in the 30 to 40 age range consistently used everyday Noun + Noun and evaluative Noun + Adjective compounds with ease, but showed markedly reduced familiarity with ritual and spiritual compounds, multi-constituent compounds, and reduplicated behavioural compounds. The most prevalent types are therefore also the most widely shared across generations, while the least prevalent are concentrated among older speakers with high traditional register competence. This pattern indicates that the frequency of a compound type in

Mambwe reflects not only its structural productivity but also the breadth of its use in everyday interaction, and that the least prevalent compound types form the stratum of the lexicon most at risk of being lost between generations.

This generational distribution has implications for language documentation and revitalisation. As Batibo (2005) warns, specialised vocabulary is among the first to be lost under language shift, and the present findings identify the specific compound types that are most vulnerable. The multi-constituent associative compounds, which encode theological and ritual meanings and were produced principally by older participants with high traditional register competence, represent a particularly urgent documentation priority. The reduplicative compounds, while semantically transparent, also showed reduced familiarity among younger speakers, suggesting that even transparent schemas may be lost if the contexts in which they are used decline.

The convergence of frequency and transparency in the Noun + Noun schema provides empirical support for the extension of Lexical Morphology Theory to Bantu languages. However, the partial exception of the reduplicative schema, which is infrequent but transparent, suggests that the theory's predictions may require refinement to account for reduplication as a distinct process. Future research might investigate whether reduplication in other Bantu languages shows a similar pattern, and whether the frequency-transparency correlation holds more strongly for compounding than for other word-formation processes.

Conclusion

This study demonstrates that Noun + Noun compounds constitute the most prevalent type of compound word in the Mambwe language of Mbala District, Zambia. Of the twenty-eight compound forms documented, Noun + Noun compounds account for 35.7 percent, followed by Noun + Adjective and Noun + Verb compounds at 21.4 percent each, while multi-constituent associative and reduplicative compounds account for 10.7 percent each. The nominal schemas together account for 67.9 percent of the corpus, confirming that Mambwe compounding is predominantly nominal. Findings further reveal that the most frequent type is also the most transparent, with eight of the ten Noun + Noun compounds being endocentric, whereas the multi-constituent associative type is entirely exocentric. Interpreted through Lexical Morphology Theory, this alignment of frequency and transparency identifies Noun + Noun compounding as productive Level II word formation that continues to expand the Mambwe lexicon, a conclusion reinforced by older speakers' coinage of new compounds for introduced referents.

The research underscores that the documentation of Mambwe must give deliberate attention to compound vocabulary, and in particular to its less prevalent types. Language documentation initiatives should prioritise the recording of multi-constituent, ritual, and spiritual compounds, together with the cultural knowledge required to interpret them, while the older speakers who hold this knowledge remain available. Curriculum developers and teachers of Mambwe should present the productive Noun + Noun and Noun + Adjective schemas as patterns for creating words rather than as fixed lists, so that learners develop the capacity to form new compounds. Dictionary makers should record compound forms as separate entries with their full meanings, since the meanings of exocentric compounds cannot be predicted from their constituents.

Future research should test these findings against a larger corpus and a larger sample of speakers, since the present corpus of twenty-eight forms was elicited from ten participants and its proportions should be treated as indicative rather than definitive. Studies based on naturally occurring speech, including ceremonial discourse, would establish whether the distribution observed in elicitation holds in everyday and ritual use. Comparative research applying the same elicitation procedure to the closely related ciLungu and ciNamwanga languages would establish whether the dominance of Noun + Noun compounding is shared across the languages of Northern Province, and would help resolve the contrast with Bemba noted above. Finally, longitudinal research tracking the same compound inventory across generations would show whether the less prevalent compound types are being lost over time.

References

1. Basciano, B., Kula, N., & Melloni, C. (2011). Modes of compounding in Bantu, Romance and Chinese. *Rivista di Linguistica, 23*(2), 203-249.
2. Batibo, H. M. (2005). *Language decline and death in Africa: Causes, consequences and challenges*. Multilingual Matters.
3. Bloomfield, L. (1933). *Language*. Henry Holt and Company.
4. Booij, G. (2012). *The grammar of words: An introduction to linguistic morphology* (3rd ed.). Oxford University Press.
5. Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
6. Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
7. Guevara, E., & Scalise, S. (2009). Searching for universals in compounding. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of language today* (pp. 101-128). Springer.
8. Guthrie, M. (1948). *The classification of the Bantu languages*. Oxford University Press.
9. Hay, J., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences, 9*(7), 342-348. https://doi.org/10.1016/j.tics.2005.05.002

10. Katamba, F. (1993). *Morphology*. Macmillan.
11. Katamba, F. (2003). Bantu nominal morphology. In D. Nurse & G. Philippson (Eds.), *The Bantu languages* (pp. 103-120). Routledge.
12. Kiparsky, P. (1982). Lexical morphology and phonology. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 3-91). Hanshin Publishing.
13. Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners* (5th ed.). SAGE Publications.
14. Lieber, R. (2009). *Introducing morphology*. Cambridge University Press.
15. Nurse, D., & Philippson, G. (Eds.). (2003). *The Bantu languages*. Routledge.
16. Okolie, U.C., Lubbungu, J., Osuji, C.U., Idike, I.M., Ugwuda, S.O., Ehiobuche, C., & Siame, P. (2025). Career-related teacher support and decision to choose teaching as a career: A three-wave study of Nigerian and Zambian university students. *International Journal for Educational and Vocational Guidance*, 25 (2), 633-661
17. Scalise, S., & Vogel, I. (Eds.). (2010). *Cross-disciplinary issues in compounding*. John Benjamins.
18. Siame, P., & Banda, F. (2024). A morphological analysis of micro-variations on verbal extensions in ciLungu, ciMambwe and ciNamwanga languages. *Comparative Linguistics Translation and Literary Studies, 1*(1), 18-34. https://doi.org/10.70036/cltls.v1i1.3
19. Siame, P., Amoakohene, B., & Kapau, H.M. (2023). Phonological Processes Governing Borrowing from English to Lungu, Mambwe, and Namwanga Languages. *Daengku: Journal of Humanities and Social Sciences Innovation*, 3 (5), 821-830. https://doi.org/10.35877/454RI.daengku1765
20. Siame, P. & Lubbungu, J. (2023). Decolonisation: Enhancing equity in the management of language education in Secondary Schools in Kabwe District. *Kwame Nkrumah University Multi-Disciplinary Journal - Zambia*, 1(1), 101-116. https://accessjournal.nkrumah.edu.zm/index.php/knuj/article/view/16
21. Siame, P. (2022). Potholes in the teaching of Zambian languages in secondary schools: A case of Bemba language. *EduLine: Journal of Education and Learning Innovation*, 2 (4), 548-560.
22. Siame, P. (2023). A morphological analysis of tense in Mambwe language. *Daengku: Journal of Humanities and Social Sciences Innovation, 3*(2), 270-278. https://doi.org/10.35877/454RI.daengku1605
23. Siame, P., Chisenga, R.P.A., Kangwa, K.N., Kapau, H.M., & Amoakohene, B. (2025). Constructing legitimacy through language: A critical discourse analysis of Zambian Parliamentary debates. *IKR Journal of Arts, Humanities and Social Sciences (IKRJAHSS)*, 1 (4), 232-243
24. Zimba, S. (2014). The grammar of the Tumbuka compound noun: A case study of Tumbuka compound nouns of the Lundazi district in the Eastern Province of Zambia. *Journal of Education Research and Behavioral Sciences, 3*(1), 35-44.