International Research Journal of Language, Literature and Culture
ISSN: 3143-0384 | Volume 1, Issue 4, 2026
www.aytinpublications.com

Research Article

Categorisation of Compound Words in the Mambwe Language of Mbala District, Zambia

Norah Nampasa[1], Pethias Siame[2]*

[1,2]*Department of Literature and Languages, Kwame Nkrumah University, Zambia*

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

Abstract

This study categorises the compound words of the Mambwe language of Mbala District, Zambia, a Bantu language classified as zone M.15 whose compounding system had not previously been documented. It reports on the second objective of a wider study of Mambwe compound morphology, the first of which established the prevalent structural types. Grounded in Lexical Morphology Theory, with Construction Grammar in a supporting role, 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 resulting corpus of twenty-eight compound forms was categorised along three dimensions: semantic headedness, semantic domain and functional register. Eighteen compounds (64.3 percent) are endocentric and ten (35.7 percent) exocentric. Five semantic categories were identified: descriptive (35.7 percent), ritual and spiritual (17.9 percent), behavioural (17.9 percent), kinship (14.3 percent) and social hierarchy (14.3 percent) compounds. Exocentricity is unevenly distributed: four of the five ritual and spiritual compounds are exocentric, whereas all kinship and behavioural compounds are endocentric. Three functional registers were identified: ceremonial and ritual use, daily interaction, and moral evaluation and identity marking. The distribution aligns with the prediction of Lexical Morphology Theory that productive, transparent compounds predominate while opaque compounds are stored as conventional items, and it shows that exocentricity clusters in the domain most dependent on cultural knowledge. The study concludes that headedness in Mambwe cannot always be read off from form alone but depends on cultural knowledge, and recommends that the culturally opaque ritual compounds be documented with their interpretations while older speakers remain available.

Keywords: Compound words, Mambwe language, categorisation, endocentric and exocentric compounds, semantic domains, Lexical Morphology Theory, Zambia.

Article Information

Received: 01 November 2026

Accepted: 13 November 2026

Published: 15 November 2026

Cite this article as:

Norah Nampasa, Pethias Siame. Categorisation of Compound Words in the Mambwe Language of Mbala District, Zambia. International Research Journal of Language, Literature and Culture. 2026; 1(4): 01-10.

Copyright:©2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Compound words do not form an undifferentiated list. Within any language, they fall into categories according to the way their meaning relates to their parts, the areas of experience they name, and the social settings in which they are used. Categorising compounds is therefore a necessary step between identifying them and explaining how they are formed, since categories reveal the principles by which speakers organise compound vocabulary. The Mambwe language, classified under Guthrie's (1948) zone M.15, is spoken in Mbala and Senga districts of Northern Province, Zambia, and in parts of south-western Tanzania (Siame, 2023). Like other Bantu languages, it has an extensive noun class system and agglutinative morphology, and much of the cultural knowledge of the Mambwe people, concerning kinship, social rank, moral conduct and the spiritual world, is carried in its compound vocabulary.

Research on Bantu morphology has concentrated on the verb, and nominal compounding has received comparatively little attention (Nurse & Philippson, 2003). Where Bantu compounds have been classified, the categories used have usually been structural, describing the word classes of the constituents, and have treated headedness as a property that can be read directly from form (Katamba, 2003). Research on Mambwe has followed the wider pattern: existing studies have examined tense and verbal extensions (Siame, 2023; Siame & Banda, 2024), and no study has established how Mambwe compound words may be categorised. This omission matters theoretically, because categories that work for European languages may not capture the organisation of a Bantu lexicon, and practically, because the categories of compounds most closely tied to cultural knowledge are the ones most likely to be lost as Mambwe gives way to English and Bemba in many domains, a vulnerability that Batibo (2005) identifies as characteristic of specialised vocabulary across African languages.

This article categorises compound words in the Mambwe language. The study addresses the research question: how can compound words be categorised in the Mambwe language? It categorises the corpus of twenty-eight compound forms along three dimensions, namely semantic headedness, semantic domain and functional register, and interprets the relationships between them through Lexical Morphology Theory. In doing so, it provides a categorisation framework for Mambwe compounding and contributes data from an under-described language to the comparative study of Bantu compound classification.

Literature Review

The classification of compounds has a long history in general linguistics. Bloomfield (1933) introduced the distinction between endocentric compounds, whose meaning is headed by one of their constituents, as in a blackbird being a kind of bird, and exocentric compounds, whose referent lies outside their constituents. Scalise and Bisetto (2009) develop this into a cross-linguistic classification in which compounds are first divided by the grammatical relation between their constituents into subordinate, attributive-appositive and coordinate classes, each of which may then be endocentric or exocentric. Lieber (2009) categorises compounds according to the semantic relationship between their constituents, and Scalise and Vogel (2010) bring together typological, psycholinguistic and semantic perspectives on compounding. These frameworks provide the vocabulary for categorising compounds in any language, but they were developed largely on Indo-European data, in which headedness can usually be established from form, and they leave open how categorisation proceeds in languages where much compound meaning rests on cultural convention.

African scholarship has shown that Bantu compounds require categories suited to their structure. Katamba (2003), drawing on Luganda, combines structural criteria based on the morphological relationship between constituents with semantic criteria reflecting the organisation of cultural knowledge, and argues that categorisation frameworks imported from European languages are inadequate for Bantu data. His account nevertheless treats endocentricity as a structural property to be read off the compound. Basciano et al. (2011), comparing Bemba with Italian and Mandarin Chinese, establish that Bantu nominal compounds are characteristically left-headed, a finding that predicts that most compounds in a Bantu language will be endocentric with the head in first position. Nurse and Philippson (2003) note that nominal compounding generates terms for kinship, social roles and material culture across the Bantu family, implying that semantic domain is a natural basis for categorisation. Batibo (2005) adds that the specialised and culturally dependent vocabulary of African languages is the most vulnerable to loss under language shift, which gives categorisation by domain and register a practical as well as a theoretical value, since a categorisation that identifies which domains are most culturally dependent also identifies where documentation effort is most urgently needed.

Within Zambia, compounding has been studied most fully for Cinyanja. Njobvu and Mambwe (2022), working within Distributed Morphology, show that the meanings of Cinyanja compounds such as cipha-dzuwa, literally 'killer of the sun' but conventionally 'a beautiful girl', are established through cultural association rather than derived from their parts, and Zimba (2014) describes the grammar of

the compound noun in Tumbuka. This work demonstrates that the meaning of many Zambian compounds cannot be read from their structure, and it therefore provides the closest regional precedent for the present study's finding that headedness in Mambwe cannot always be established from form alone. For Mambwe itself, the work of Siame (2023) and Siame and Banda (2024) establishes the noun class and agreement patterns against which compounds can be described, but it is oriented towards the verb and offers no account of compounding. No study has categorised Mambwe compounds by headedness, semantic domain or functional register, and none has asked whether the endocentric and exocentric categories can be assigned on formal grounds alone in this language. This is the gap the present study addresses.

Theoretical Framework

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 with its own morphological and phonological rules. Although Kiparsky's formulation was concerned principally with derivation, the framework was 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. On this extended version of the theory, compounds formed productively at Level II are regular and semantically transparent, whereas compounds whose meanings cannot be recovered from their constituents are stored as conventionalised items in the manner characteristic of Level I.

The principle most directly relevant to categorisation is that of affixation and rule ordering, together with the related principle of category inheritance. A compound formed by a productive rule inherits its lexical and semantic category from its internal head; it is endocentric. A compound that inherits no category from within must have its category assigned by convention; it is exocentric, and on the extended theory it is a stored rather than a generated item. The theory therefore predicts that endocentric compounds will predominate in a productive compounding system and that exocentric compounds will be concentrated in those areas of the lexicon where meanings are conventional rather than compositional; this is the prediction against which the distribution reported in Section 5.1 is assessed.

Applied to the present study, Lexical Morphology Theory supplied the primary dimension of categorisation, headedness, and the prediction against which its distribution was assessed. Construction Grammar (Goldberg, 2006) was used in a supporting role. It treats form-meaning pairings at every level of size as conventional constructions, and so accounts for the culturally specific meanings of exocentric forms, which a purely rule-based account leaves unexplained, including the culturally established meanings that Njobvu and Mambwe (2022) document for Cinyanja compounds. It also motivated the second and third dimensions of categorisation, semantic domain and functional register, since on this view the meaning of a construction includes knowledge of the contexts in which it is used.

Methodology

This study adopted a qualitative approach within an interpretivist paradigm, which was suited to understanding how Mambwe speakers construct and interpret compound words within their cultural context (Creswell & Creswell, 2023). An exploratory research design was employed because the compounding system of Mambwe had not previously been documented, and the study sought to describe it in its natural setting without manipulating any variables. The exploratory design was appropriate because no prior categorisation framework existed for Mambwe compounds, and the study therefore aimed to generate categories grounded in speakers' own usage rather than to test pre-existing hypotheses (Creswell & Creswell, 2023). The interpretivist paradigm further required that meaning be treated as socially constructed, so that participants' explanations of compound meanings and contexts were treated as primary data rather than as background information (Kumar, 2019).

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). Purposive sampling was chosen because the study required participants who could both produce compound forms and explain their cultural meanings, a criterion that random sampling could not guarantee; snowball sampling was then used to reach further specialists, such as traditional healers and cultural custodians, who are not easily identified through formal community structures (Kumar, 2019). The age range of 30 to 65 years was set deliberately: it includes older speakers who retain the traditional register and younger adults who use compounds in everyday interaction, allowing the study to compare familiarity across generations (Batibo, 2005). 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. Competence ratings were assigned by the first author in consultation with two senior participants, P01 and P08, on the basis of each participant's demonstrated ability to produce and explain ritual and ceremonial vocabulary during elicitation (Kumar, 2019). 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. This procedure follows the principle that saturation, rather than a predetermined numerical target, should determine sample size in qualitative research (Creswell & Creswell, 2023; Kumar, 2019).

Data were collected through four instruments, refined through a pilot study with three speakers. The pilot study was conducted with three Mambwe speakers drawn from the same population but not included in the main sample, and it was used to test the clarity of interview questions, the coverage of the semantic domains in the elicitation matrix, and the adequacy of the recording procedure; ambiguous items identified during the pilot were reworded before the main fieldwork began (Kumar, 2019). A semi-structured interview guide explored the use and meaning of compound words in daily and cultural life. 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, explain their parts and meanings, and state where and by whom they are used. The matrix was organised by semantic domain because nominal compounding in Bantu languages is known to generate vocabulary for kinship, social roles and material culture, making domain a natural basis for elicitation as well as for later categorisation (Nurse & Philippson, 2003; Katamba, 2003). Two focus group discussions, one with participants of different ages and one with cultural experts, validated the elicited forms, their meanings and their contexts of use, and participant observation at community gatherings documented compound use in natural conversation. Sessions were audio-recorded and supplemented by field notes.

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 compounds was assembled. In the second phase, each compound was segmented into its constituent morphemes and assigned to a structural schema. Segmentation followed the noun class and agreement patterns described for Mambwe by Siame (2023) and Siame and Banda (2024), which provided the grammatical reference against which constituent boundaries and heads could be identified. In the third phase, the corpus was categorised along three dimensions. For headedness, a compound was classified as endocentric if one of its constituents determined the category of its referent, and as exocentric if its referent could not be identified from either constituent without cultural or pragmatic knowledge; this operational definition follows Bloomfield's (1933) distinction and Scalise and Bisetto's (2009) cross-linguistic classification, adapted to a language in which headedness cannot always be read from form (Katamba, 2003; Njobvu & Mambwe, 2022). For semantic domain, each compound was assigned to the domain that formed its semantic centre of gravity, recognising that some forms belong to more than one; the domains were derived from the elicitation matrix and refined inductively as new forms were analysed (Nurse & Philippson, 2003). For functional register, compounds were grouped according to participants' reports of the contexts in which they are used; register was included as a third dimension because Construction Grammar treats knowledge of context as part of the meaning of a construction (Goldberg, 2006). The broad domain assignments were reviewed with two senior participants, P01 and P08, who confirmed them, and preliminary analyses were returned to all participants for member checking, a procedure that strengthens the credibility of qualitative interpretation by allowing participants to confirm or correct the researcher's account (Creswell & Creswell, 2023).

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). 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. These procedures were adopted because the research involved culturally sensitive knowledge held by identifiable specialists, and the risk of harm arose less from physical intrusion than from the inappropriate disclosure of restricted ritual vocabulary (Creswell & Creswell, 2023).

Results and Discussion

This section presents the categorisation of Mambwe compound words. Compounds are cited in the practical

orthography of Mambwe, with constituents separated by a plus sign and English glosses given in single quotation marks.

Categorisation of Compound Words

The corpus of twenty-eight compound forms was categorised by semantic headedness and by semantic domain, as summarised in Table 1. Eighteen forms (64.3 percent) are endocentric, and ten (35.7 percent) are exocentric. Five semantic categories were identified. Descriptive compounds form the largest category, with ten forms (35.7 percent), followed by ritual and spiritual compounds and behavioural compounds with five forms each (17.9 percent), and by kinship compounds and social hierarchy compounds with four forms each (14.3 percent). The compounds belonging to each category are listed in Table 1.

The analysis in this section proceeded in two steps. First, each compound was assigned to a semantic category on the basis of the domain that formed its semantic centre of gravity, following the elicitation matrix and the confirmations of participants P01 and P08. Second, each compound was classified as endocentric or exocentric using the operational definition given in Section 4: a compound is endocentric where one constituent determines the category of the referent, and exocentric where the referent cannot be identified from either constituent without cultural or pragmatic knowledge (Bloomfield, 1933; Scalise & Bisetto, 2009). The two classifications were then cross-tabulated, and the resulting distribution is the basis of the discussion that follows. Where a compound could plausibly be assigned to more than one domain, it was counted once, under the domain that participants themselves identified as primary; this affected two forms, Musamaliya-musuma and Kwikala-waulanda, which are discussed in Section 5.3.

Table 1. *Distribution of Mambwe Compounds by Semantic Category and Headedness (N = 28)*

Semantic category	Endocentric	Exocentric	Total	%
Kinship	4	0	4	14.3
Social hierarchy	2	2	4	14.3
Ritual and spiritual	1	4	5	17.9
Descriptive	6	4	10	35.7
Behavioural	5	0	5	17.9
Total	**18**	**10**	**28**	**100.0**

The endocentric results show that the referent of the compound is determined by one of its constituents, while the exocentric results reveal that the referent lies outside its constituents. Each compound is counted once, under the category that forms its semantic centre of gravity. Percentages do not sum to exactly 100 because of rounding.

The distribution in Table 1 shows three patterns that require explanation. First, endocentric compounds outnumber exocentric compounds by roughly two to one, which is consistent with the prediction of Lexical Morphology Theory that a productive compounding system will be dominated by forms generated by transparent rules (Kiparsky, 1982; Booij, 2012). Second, exocentricity is not spread evenly across the corpus but is concentrated in two categories, ritual and spiritual (four of five forms) and descriptive (four of ten forms), and is absent from kinship and behavioural compounds.

Third, the category with the largest number of forms, descriptive compounds, is also internally divided, with six endocentric and four exocentric members. These patterns indicate that headedness in Mambwe is not a property of the compound system as a whole but varies systematically with the area of experience the compound names, a point developed in Sections 5.2 to 5.5.

Table 2. *Mambwe Compounds by Semantic Category*

Semantic category	Compounds (Glossed)
Kinship	Taata-vyala ‘father-in-law’; Nyoko-vyala ‘mother-in-law’; Umwana-mwanaci ‘baby girl’; Umwana-monsi ‘baby boy’
Social hierarchy	Mwine-muzi ‘village headman’; Mwene-mukulu ‘great chief, lord’; Musamaliya-musuma ‘Good Samaritan’; Kacema-muzuma ‘Good Shepherd’
Ritual and spiritual	Mupashi-wamazilo ‘Holy Spirit’; Mupasi-wamuzilo ‘spirit of sacred breath’; Indya-muluwa ‘spiritual nourishment’; Yemba-wachilolwa ‘sea of glass’; Ifwa-linowalya ‘death that eats’
Descriptive	Muntu-musuma ‘good person’; Mwana-mwipe ‘bad child’; Imbuzi-yalwango ‘hungry goat’; Cimela-cinono ‘small growing thing’; Icintu-wingi ‘something important’; Simpu-lwilo ‘fast runner’; Ikanda-ngombe ‘heavy rain’; Kaa-kolwe ‘small agile monkey; agile person’; Mama-waluse ‘mother of mercy’; Ciumbwa-cacoleze ‘spoiled or corrupted thing’
Behavioural	Ukutunda-tunda ‘frequent urination’; Ukukolwa-kolwa ‘being perpetually drunk’; Ukumwenjwa-manzi ‘frequent drinking of water’; Kwikala-waulanda ‘living in poverty’; Zuwa-zuwa ‘very fast’

Endocentric and Exocentric Compounds

The predominance of endocentric compounds confirms that the Mambwe compound system is primarily head-internal. In endocentric compounds, the head is usually the first constituent, consistent with the left-headed pattern Basciano et al. (2011) report for Bemba. In Mwine-muzi (mwine 'owner' + muzi 'village') 'village headman', the head mwine determines that the compound refers to a person. In Taata-vyala 'father-in-law' and Nyoko-vyala 'mother-in-law', the parental terms taata and nyoko define the person-denoting category of the whole, and in Umwana-mwanaci 'baby girl' and Umwana-monsi 'baby boy' the head umwana 'child' is narrowed by the second constituent. Muntu-musuma 'good person' and Mwana-mwipe 'bad child' are likewise headed by the nouns muntu and mwana.

This left-headed pattern is the one Basciano et al. (2011) establish for Bantu nominal compounds generally, and the Mambwe data therefore align with the wider Bantu family rather than constituting an exception. The finding also supports Katamba's (2003) observation that Bantu nominal morphology is organised around the noun class system, since in each of these examples the class prefix of the first constituent determines the agreement pattern of the whole compound (Siame, 2023).

Exocentric compounds, by contrast, denote referents that neither constituent names. Table 3 sets out the literal meanings of their constituents against their actual referents. Ikanda-ngombe, literally 'cattle hide', denotes heavy, torrential rain. Kaa-kolwe, a small agile monkey, is applied figuratively to a quick or agile person, shifting its referent from an animal to a human being. Mupashi-wamazilo (mupashi 'spirit, breath' + wa- + mazilo 'holiness') 'Holy Spirit' and Yemba-wachilolwa 'sea of glass' take their meanings from religious and cosmological knowledge that the constituents themselves do not supply.

These examples resemble the Cinyanja compounds described by Njobvu and Mambwe (2022), such as ciphadzuwa 'killer of the sun', whose meaning 'a beautiful girl' is established through cultural association rather than derived from its parts. The parallel is significant because it shows that the phenomenon is not idiosyncratic to Mambwe but is characteristic of compounding in Zambian languages, and that any categorisation framework for these languages must therefore provide for conventionalised meaning alongside compositional meaning.

Table 3. *Exocentric Mambwe Compounds: Literal versus Actual Referents*

Compound	Literal constituent meanings	Actual referent
Ikanda-ngombe	hide + cattle	Heavy or torrential rain
Kaa-kolwe	monkey + small, agile	Quick or agile person (figurative)
Yemba-wachilolwa	water + shining	Sea or body of water with a luminous surface
Mupashi-wamazilo	spirit, breath + holiness	Holy Spirit
Mupasi-wamuzilo	spirit + sacred breath	Spirit of sacred breath
Indya-muluwa	food, eating + spirit	Spiritual nourishment
Mama-waluse	mother + mercy	A merciful woman
Musamaliya-musuma	caring one + good	Good Samaritan
Kacema-muzuma	shepherd + good	Good Shepherd
Ciumbwa-cacoleze	thing + corrupted	A spoiled or morally corrupted entity

This distribution is the one Lexical Morphology Theory predicts. Endocentric compounds inherit their category from an internal head, as expected of forms generated by productive and transparent Level II rules (Booij, 2012; Kiparsky, 1982). Exocentric compounds inherit no category from within and must be assigned one by convention, so that on the extended theory they are stored items in the manner of Level I. The predominance of endocentric forms indicates that speakers favour predictable schemas for everyday reference, while exocentric constructions are reserved for culturally embedded, ceremonial and figurative meanings. The finding also qualifies Katamba's (2003) treatment of endocentricity as a structural property. For ten of the twenty-eight Mambwe compounds, headedness could not be established from form alone but only with the cultural knowledge that motivates the compound's meaning. An analysis confined to formal criteria would have recorded these forms as anomalous rather than as culturally motivated.

This qualification of Katamba (2003) is consistent with the position taken by Njobvu and Mambwe (2022), who show that the meanings of Cinyanja compounds are established through cultural association and cannot be predicted from their constituents. It also supports the supplementary role assigned to Construction Grammar in this study, since Goldberg (2006) treats culturally specific meanings as conventional constructions learned from use rather than derived by rule. The practical implication is that a categorisation of Mambwe compounds which relied on form alone would misclassify a substantial minority of the

corpus, and would obscure precisely those forms in which cultural knowledge is most densely encoded (Batibo, 2005).

Semantic Categories

Five semantic categories were identified. The categories are best understood as prototypical clusters rather than as a strict partition of the corpus, since several forms belong to more than one: Musamaliya-musuma 'Good Samaritan' has both social and behavioural dimensions, and Kwikala-waulanda 'living in poverty' lies between the behavioural and descriptive domains. Each compound was assigned to the category that forms its semantic centre of gravity. This treatment of categories as prototypical rather than discrete follows the Construction Grammar position that constructions are organised around central members with peripheral extensions (Goldberg, 2006), and it is preferable to a strict partition because it preserves the overlap that participants themselves recognised when explaining the meanings of these forms.

Kinship compounds form the most internally consistent category, and all four are endocentric. Taata-vyala and Nyoko-vyala combine a parental term with vyala, a nominalised form of the verb -vyala 'to beget, to give birth', and the schema [parental term + vyala] generates affinal kinship terms by substitution of the parental term. Kinship terminology in Mambwe is therefore morphologically generative rather than merely listed, supporting Katamba's (2003) observation that Bantu kinship vocabulary is formed by productive schemas. The pair Umwana-mwanaci 'baby girl' and Umwana-monsi 'baby boy' encodes gender within the same head-modifier pattern. The uniform endocentricity of this category is consistent with the prediction of Lexical Morphology Theory that productive schemas generate transparent forms at Level II (Booij, 2012), and with Nurse and Philippson's (2003) observation that nominal compounding across the Bantu family generates kinship terms. Because the schema is productive, it is also the category least immediately threatened by language shift, since new kinship terms can be formed by substitution rather than learned as fixed items (Batibo, 2005).

Social hierarchy compounds encode the ranks, roles, and relational positions that structure Mambwe community life. Mwine-muzi 'village headman' and Mwene-mukulu 'great chief, lord' are endocentric designations for positions of authority, whereas Musamaliya-musuma 'Good Samaritan' and Kacema-muzuma 'Good Shepherd' are exocentric, naming social and moral roles introduced through religious contact. These compounds were attested consistently across all participant groups, confirming their status as conventionalised social labels rather than descriptions created on the spot. The division within this category is instructive. The two endocentric forms name positions internal to Mambwe social organisation, whereas the two exocentric forms name roles introduced through religious contact and therefore depend on knowledge external to the indigenous social system. This pattern supports the Construction Grammar view that a construction's meaning includes knowledge of the contexts in which it is used (Goldberg, 2006), since the exocentric members of this category cannot be interpreted without knowledge of the religious tradition from which they derive.

Ritual and spiritual compounds form the most semantically opaque category in the corpus: four of the five are exocentric. Mupashi-wamazilo 'Holy Spirit', Mupasi-wamuzilo 'spirit of sacred breath', Indya-muluwa 'spiritual nourishment' and Yemba-wachilolwa 'sea of glass' map physical realities, such as breath, food and water, onto spiritual meanings, such as divine life, sustenance and purity. Only Ifwa-linowalya 'death that eats', headed by ifwa 'death', is endocentric. The concentration of exocentricity in this category shows that the most symbolically significant vocabulary is encoded in the most culturally dependent structures. This concentration is exactly what Lexical Morphology Theory predicts, since exocentric forms are stored as conventionalised items whose meanings must be learned rather than generated (Kiparsky, 1982; Booij, 2012). It also identifies this category as the most vulnerable in the lexicon, because the specialised vocabulary of African languages is the most exposed to loss under language shift (Batibo, 2005), and because the meanings of these forms cannot be reconstructed from their parts by speakers who no longer possess the associated cultural knowledge (Njobvu & Mambwe, 2022).

Descriptive compounds form the largest category, encoding physical, evaluative and qualitative characterisations of persons, animals and entities. Six are endocentric, among them Muntu-musuma 'good person', Mwana-mwipe 'bad child', Imbuzi-yalwango 'hungry goat', Cimela-cinono 'small growing thing' and Simpu-lwilo 'fast runner', in which a nominal head is qualified by an evaluative or qualitative modifier. The four exocentric descriptive compounds, Ikanda-ngombe, Kaa-kolwe, Mama-waluse 'mother of mercy, a merciful woman' and Ciumbwa-cacoleze 'a spoiled or corrupted thing', all involve figurative extension, in which a literal image is redeployed to characterise a person, an event or a moral condition. The internal division of this category is consistent with the general pattern of the corpus: the endocentric members are compositional descriptions whose meaning is recoverable from the head and modifier, while the exocentric members

depend on figurative extension of the kind Goldberg (2006) treats as a conventional construction. The presence of figurative extension within the largest category indicates that exocentricity in Mambwe is not confined to ritual vocabulary but is a general resource of the compound system, employed wherever a literal image can be redeployed to characterise a person or condition.

Behavioural compounds encode recurring patterns of action, often with moral significance, and all five are endocentric. Ukutunda-tunda 'frequent urination', Ukukolwa-kolwa 'being perpetually drunk' and Zuwa-zuwa 'very fast' use reduplication to express repeated, habitual or intensified action. Ukumwenjwa-manzi 'frequent drinking of water' combines a passive verbal form with the noun manzi 'water', so that the subject is presented as undergoing the activity rather than initiating it, and Kwikala-waulanda 'living in poverty' describes a persistent condition. Because the meaning of a reduplicated form remains recoverable from the repeated base, behavioural compounds are uniformly transparent. The uniform transparency of this category follows from the semantics of reduplication, which iconically signals repetition or intensity while preserving the meaning of the base (Katamba, 1993). In Lexical Morphology terms, these forms are generated by a productive Level II rule and inherit their category from the verbal base, which accounts for the absence of exocentricity in this category (Booij, 2012). The category is therefore the clearest example in the corpus of the transparent, rule-generated stratum of the Mambwe lexicon.

Functional and Pragmatic Categories

Beyond their semantic content, Mambwe compounds are used differently across social contexts. Participants' reports identified three functional registers, summarised in Table 4. These categories rest primarily on participant self-reports and metalinguistic commentary during elicitation, supplemented by participant observation, rather than on systematic analysis of naturally occurring discourse, and should be read with that limitation in mind.

Table 4. *Functional Registers of Mambwe Compound Words*

Register	Function	Examples
Ceremonial and ritual	Marks ceremonial speech; signals specialised cultural knowledge	Mupashi-wamazilo, Indya-muluwa, Ifwa-linowalya
Daily interaction	Reference, evaluation and social positioning in everyday talk	Mwine-muzi, Taata-vyala, Muntu-musuma, Kaa-kolwe
Moral evaluation and identity marking	Praise, criticism and moral instruction; epithets and honorifics	Mwana-mwipe, Ciumbwa-cacoleze, Mama-waluse, Simpu-lwilo, Kacema-muzuma

Compounds of the ritual and spiritual category are largely restricted to ceremonial registers. Participants consistently reported that forms such as Mupashi-wamazilo, Indya-muluwa and Ifwa-linowalya are used in religious ceremony, ancestral invocation and communal ritual rather than in casual conversation. Their use marks the shift into ceremonial speech and signals the speaker's access to specialised cultural knowledge. The participants with high traditional register competence, notably P01, P04, P06 and P08, were the principal producers of these forms, indicating that ritual compound knowledge is stratified by cultural role rather than shared evenly across the speech community. This pattern matches Batibo's (2005) observation that specialist vocabulary in African communities is held by those with cultural authority, and it is that vocabulary which is most exposed to loss. The restriction of these compounds to ceremonial contexts also has consequences for transmission. In Construction Grammar terms, the meaning of a construction is learned from its use (Goldberg, 2006), but the use from which the meanings of ritual compounds would be learned is itself confined to ceremonial settings and to a small number of authorised speakers. The channel through which this stratum of the lexicon is transmitted is therefore narrow, which reinforces Batibo's (2005) concern about the vulnerability of specialist vocabulary.

Daily interactional compounds, such as Mwine-muzi, Taata-vyala, Muntu-musuma, Mwana-mwipe and Kaa-kolwe, serve reference, evaluation and social positioning in everyday communication. They were attested across all participant groups and are core items of the Mambwe vocabulary. In Construction Grammar terms, they are not merely referential units but conventionalised schemas that carry pragmatic information about the relationship between speaker and referent (Goldberg, 2006). Their attestation across all participant groups, including the younger participants aged 30 to 40, indicates that this register is secure and is being transmitted through ordinary daily interaction rather than through specialised instruction.

The third register combines moral evaluation and identity marking. Compounds such as Mwana-mwipe, Ciumbwa-cacoleze, Kwikala-waulanda and Mama-waluse are used in praise, criticism and moral instruction, so that using them performs a social act of evaluation that draws on shared community values. Compounds such as Simpu-lwilo, Mwene-mukulu and Kacema-muzuma are used as epithets, honorifics and characterising names, through which

social identities are constructed and recognised. In this register, compound morphology serves not only to name but to position people within the moral and social order of the community. This register illustrates most clearly the Construction Grammar claim that the meaning of a construction includes knowledge of the contexts in which it is used (Goldberg, 2006), since the evaluative force of these compounds cannot be derived from their constituents alone but depends on shared community values. It also shows that compounding in Mambwe performs a social function that a purely structural classification would leave unrecorded.

Relationships among the Categories

Read together, the three dimensions of categorisation reveal a consistent pattern. Transparent, endocentric compounds dominate the kinship, behavioural and everyday descriptive categories and are used across the whole speech community. Opaque, exocentric compounds are concentrated in the ritual and spiritual category and in figurative description, and their use is restricted to ceremonial settings and to speakers with traditional register competence. This alignment of headedness, domain and register supports the extended form of Lexical Morphology Theory: the forms generated productively at Level II are those used in everyday reference, while the forms stored as conventional items are those whose meanings depend on specialised cultural knowledge. The alignment is not perfect, since the descriptive category contains both transparent and opaque members, but the overall pattern is clear and statistically marked: all ten exocentric compounds fall within the ritual and spiritual and descriptive categories, while all nine kinship and behavioural compounds are endocentric. This distribution is unlikely to have arisen by chance and indicates that headedness, domain and register are systematically related rather than independent dimensions of classification.

The pattern also shows where the limits of a purely rule-based account lie. Construction Grammar holds that the meaning of a construction is learned from its use (Goldberg, 2006), but the use from which the meanings of Mambwe ritual compounds would be learned is itself confined to ceremonial contexts. Younger participants, aged 30 to 40, showed markedly reduced familiarity with ritual and spiritual compounds, while using everyday kinship and descriptive compounds with ease. The stratum of the lexicon that is most opaque is therefore also the stratum with the narrowest channel of transmission. Because the meaning of an exocentric compound cannot be reconstructed from its parts, the loss of such a form entails the loss not only of a word but of the cultural mapping it preserves. Categorisation thus has a practical consequence: it identifies the ritual and spiritual category, and exocentric compounds generally, as the priority for documentation. This conclusion is consistent with Batibo's (2005) argument that the specialised and culturally dependent vocabulary of African languages is the most vulnerable to loss under language shift, and with Njobvu and Mambwe's (2022) demonstration that the meanings of Zambian compounds are established through cultural association. It also gives empirical support to the recommendation, made in Section 6, that documentation should record the cultural interpretations of exocentric compounds alongside their forms, since once the interpreting knowledge is lost the forms themselves become uninterpretable.

Conclusion

This study demonstrates that Mambwe compound words can be categorised along three related dimensions: semantic headedness, semantic domain and functional register. Of the twenty-eight compounds documented, eighteen (64.3 percent) are endocentric and ten (35.7 percent) exocentric. Five semantic categories were identified: descriptive compounds (35.7 percent), ritual and spiritual compounds (17.9 percent), behavioural compounds (17.9 percent), kinship compounds (14.3 percent) and social hierarchy compounds (14.3 percent). Three functional registers were identified: ceremonial and ritual use, daily interaction, and moral evaluation and identity marking. The dimensions are closely related. Exocentricity is concentrated in the ritual and spiritual category, where four of the five compounds are exocentric, and is absent from the kinship and behavioural categories, while ritual compounds are confined to ceremonial registers and to speakers with traditional register competence. Interpreted through Lexical Morphology Theory, this pattern distinguishes transparent compounds generated at Level II for everyday reference from opaque compounds stored as conventional items for culturally specialised meanings. The study further shows that in Mambwe the headedness of a substantial minority of compounds cannot be established from form alone, so that categorisation must integrate cultural interpretation with structural description.

The research underscores that the documentation of Mambwe must record not only compound forms but the categories to which they belong and the knowledge needed to interpret them. Language documentation initiatives should give priority to exocentric and ritual compounds, recording their literal and conventional meanings and their contexts of use while the older speakers who hold this knowledge remain available. Dictionary makers should enter compounds as separate headwords, indicate their semantic domain and register, and give the cultural explanation of exocentric forms, whose meanings cannot

be predicted from their parts. Curriculum developers and teachers of Mambwe should present the kinship, descriptive and behavioural schemas as productive patterns that learners can use to form new words, and should introduce learners to culturally specialised compounds together with the knowledge that gives them meaning.

Future research should test these categories against a larger corpus and sample, 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 of naturally occurring speech, including ceremonial discourse, would establish whether the functional registers reported by participants correspond to actual patterns of use. Comparative research applying the same categorisation to ciLungu and ciNamwanga would show whether the alignment of exocentricity with the ritual domain is shared across the languages of Northern Province. Finally, longitudinal research tracking the same compounds across generations would show whether the ritual and spiritual category is 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. Guthrie, M. (1948). *The classification of the Bantu languages.* Oxford University Press.
8. Katamba, F. (1993). *Morphology*. Macmillan.
9. Katamba, F. (2003). Bantu nominal morphology. In D. Nurse & G. Philippson (Eds.), *The Bantu languages* (pp. 103–120). Routledge.
10. Kiparsky, P. (1982). Lexical morphology and phonology. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 3–91). Hanshin Publishing.
11. Kumar, R. (2019). *Research methodology:* A step-by-step guide for beginners (5th ed.). SAGE Publications.
12. Lieber, R. (2009). *Introducing morphology.* Cambridge University Press.
13. Njobvu, N., & Mambwe, K. (2022). Unravelling the complexity of Cinyanja compound words: A distributed morphology approach. *Journal of Law and Social Sciences,* 6(2), 26–38.
14. Nurse, D., & Philippson, G. (Eds.). (2003). *The Bantu languages*. Routledge.
15. Scalise, S., & Bisetto, A. (2009). The classification of compounds. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp. 34–53). Oxford University Press.
16. Scalise, S., & Vogel, I. (Eds.). (2010). *Cross-disciplinary issues in compounding.* John Benjamins.
17. 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
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. 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.