

Reconstruction of Proto-Chagga Forms: A Comparative Approach

Lea Mpobela¹ & Martha J. Mmary¹

¹ Department of Foreign Languages and Linguistics, The Saint Augustine University of Tanzania, Tanzania

Correspondence: Lea Mpobela, Department of Foreign Languages and Linguistics, The Saint Augustine University of Tanzania, Tanzania.

Received: May 20, 2026

Accepted: June 25, 2026

Published: June 30, 2026

doi:10.65343/jlsp.v2i1.113

URL: <https://doi.org/10.65343/jlsp.v2i1.113>

Abstract

This study examines the reconstruction of selected Proto-Chagga lexical roots using the comparative method in historical linguistics. Data were collected from seven Chagga languages, including Kimochi, Kiuru, Kivunjo, Kikahe, Kiwoso, Kirombo, and Kimashami. Data were collected from 14 informants, including two native speakers of each of the languages. The informants were selected using convenience sampling. The analysis followed a systematic four-step reconstruction procedure: assembling cognate sets, establishing sound correspondences, reconstructing proto-forms, and identifying innovation. The fifth step of tabulating the innovations to determine subgroup relationships was done generally for all the words. Reconstruction was guided by two major principles: the majority principle and phonological or conventional wisdom/phonetic plausibility. The findings reveal regular sound correspondences across the varieties and demonstrate how processes such as glide deletion, vowel deletion, vowel lengthening, coalescence, palatalization, and lexical replacement have contributed to the diversification of the languages. The study confirms the historical relationship among the Chagga varieties, which makes them belong to the same proto-Chagga ancestral language.

Keywords: Proto-Chagga, comparative method, language reconstruction, sound correspondences, historical linguistics

1. Introduction

Language reconstruction is the process of systematically reconstructing the features of an ancestral language that is no longer directly attested. It relies on linguistic evidence from related languages and applies comparative and internal reconstruction methods. The goal of this process is to determine the phonological, morphological, syntactic, and lexical characteristics of the proto-language, which serves as the common ancestor of a language family (Campbell, 2013). The practice of language reconstruction dates back to the 19th century, with scholars like Franz Bopp and August Schleicher pioneering the field (Hock & Joseph, 2019). Their work led to the establishment of the comparative method, which allows linguists to infer the features of unattested proto-languages by comparing systematically related words and sounds in descendant languages. Through this method, linguists have successfully reconstructed various proto-languages, such as Proto-Indo-European, Proto-Bantu, and Proto-Austronesian (Trask, 2000).

Language reconstruction is guided by two approaches: the comparative historical linguistic approach and internal reconstruction. The comparative historical linguistic approach assumes that languages change systematically over time. By comparing related languages and identifying regular sound correspondences, linguists can reconstruct proto-forms (Fox, 1995). Scholars such as Guthrie (1967), Meeussen (1967), Nurse and Philippson (2003), and Philippson (2006) have demonstrated that the comparative method constitutes an effective tool for reconstructing Proto-Bantu phonology, morphology, and lexicon. These reconstructions have not only clarified internal subgrouping within Bantu but have also contributed to broader discussions concerning migration patterns, cultural history, and contact among African communities (Ehret, 2001). Internal reconstruction analyses irregularities within a single language to infer earlier linguistic stages. It is beneficial when external comparative data is limited (Campbell, 2013). Despite these advances, many Bantu subgroups and language clusters remain underexplored from a historical-comparative perspective, particularly at lower genealogical levels such as dialect groups.

Reconstruction at the level of Proto-Bantu alone, however, does not suffice to explain the full range of variation found among present-day Bantu languages. As languages continue to diverge following the breakup of Proto-Bantu, innovations emerge within smaller subgroups. For this reason, historical linguistics increasingly recognizes the importance of reconstructing intermediate proto-languages that capture developments occurring between the Proto-Bantu stage and modern languages (Nurse, 1988; Schadeberg, 2009). Such intermediate

reconstructions allow linguists to identify shared innovations, trace patterns of internal diversification, and establish more precise subgrouping relationships.

Therefore, this study aims to address these problems by analyzing sound changes in selected words from seven Chagga varieties using the five-step reconstruction method. This is done by reconstructing the Proto-forms of selected Chagga terms through the systematic application of the comparative method. Specifically, the researchers analyse regular sound correspondences among selected Chagga varieties as a basis for reconstructing Proto-Chagga phonological forms; and reconstruct selected Proto-Chagga lexical items by comparing cognate sets across Chagga varieties and determining their historical development.

2. Literature Review

Historical and comparative linguistics rests upon methodological principles that have been refined over more than two centuries of scholarship. At the core of this discipline lies the Comparative Method, which Campbell (2013) and Trask (2015) identify as the most reliable tool for reconstructing proto-languages. Both scholars converge on the position that systematic comparison of cognate forms across related languages enables linguists to establish regular sound correspondences, thereby facilitating the reconstruction of phonological, morphological, and lexical systems even in the absence of written attestation. Campbell (2013) particularly emphasizes that the strength of the Comparative Method derives from its ability to uncover historical regularities that would otherwise remain inaccessible. Trask (2015) extends this argument by demonstrating how the method's predictive power allows linguists to test reconstructive hypotheses against empirical data from descendant languages.

Complementing the Comparative Method, Hock and Joseph (2009) argue that internal reconstruction serves as an indispensable auxiliary technique for uncovering earlier linguistic stages. This is evidenced by Mpobela (2022), who traces the two lost vowels through what she refers to as the scars left by the lost vowels. She presents the attested synchronic morphophonological processes of Bantu spirantization, affrication, and palatalization, which are inherently caused by the two vowels (i, u) which are no longer attested in the language. Hock and Joseph demonstrate that morphophonemic alternations, suppletive forms, and irregular paradigms frequently preserve fossils of earlier systematic patterns. Nevertheless, they acknowledge that internal reconstruction operates within narrower limits than comparative analysis, as it cannot by itself distinguish between inherited archaisms and language-internal innovations without external comparative evidence, as done by Mpobela (2022) in Runyambo.

The relationship between these two methodological approaches has generated considerable debate. While Hock and Joseph (2009) present them as complementary, Campbell (2013) maintains that the Comparative Method takes logical and procedural precedence, with internal reconstruction serving primarily to refine or extend hypotheses generated through comparison. The application of historical-comparative methods to African languages has required scholars to adapt and refine established approaches to accommodate the continent's distinctive linguistic landscape. Ehret (2001) argues forcefully that historical reconstruction assumes particular importance in African contexts due to the scarcity of early written documentation for most language communities. In the absence of textual records, linguistic evidence constitutes the primary source for reconstructing Africa's linguistic and population history, including patterns of migration, contact, and cultural change over millennia.

Nurse and Philippson (2003) adopt a more circumscribed approach, focusing specifically on the Bantu language family while acknowledging its broader implications for African historical linguistics. They demonstrate that comparative linguistic data have been central to understanding the Bantu expansion, one of the most significant demographic and linguistic events in African prehistory. By synthesizing phonological, morphological, and lexical evidence across hundreds of Bantu languages, Nurse and Philippson establish a framework for analysing genetic relationships that has become standard in the field. Their work illustrates both the explanatory power of comparative methods and the practical challenges of applying them at a continental scale.

The reconstruction of Proto-Bantu stands as one of the most significant achievements in African historical linguistics, establishing a methodological benchmark for subsequent research. Meeussen (1967) provided the foundational synthesis, demonstrating through systematic comparison that core aspects of Bantu grammar, particularly noun class systems and verbal morphology, could be reconstructed with considerable confidence. His work established a rigorous methodological model that integrated phonological, morphological, and lexical evidence within a coherent analytical framework. Meeussen's Proto-Bantu reconstructions continue to serve as the primary reference point for comparative studies across the family, testifying to the durability of his contributions.

Building on this foundation, Schadeberg (2003) refined Proto-Bantu phonological reconstructions, while emphasizing the centrality of regular sound correspondences for explaining patterns of Bantu diversification. Schadeberg's work demonstrates how attention to phonetic detail and conditioning environments can resolve apparent irregularities and strengthen the empirical basis for reconstruction. Similarly, Nurse and Philippson (2003) expanded the reconstructed lexicon while integrating phonological and morphological evidence into a more comprehensive account of Proto-Bantu structure and its subsequent development across the family.

Despite these achievements, the Proto-Bantu enterprise has not escaped criticism. The reconstructed system, while internally coherent, may represent an idealized abstraction that smooths over variation and alternative possibilities. Different scholars have proposed competing reconstructions for particular features, reflecting ongoing debates about the interpretation of comparative data. Moreover, the reliance on a relatively small set of widely distributed languages for core reconstructions may bias the proto-language toward the features of particular subgroups, potentially obscuring the original diversity of Proto-Bantu itself.

More fundamentally for the present study, Proto-Bantu reconstruction operates at a level of generality that cannot capture developments specific to individual subgroups. As Nurse (1988) and Schadeberg (2009) argue, many significant innovations occurred after the breakup of Proto-Bantu and can only be adequately understood through reconstruction at intermediate levels. Proto-Bantu provides a necessary baseline, but it does not and cannot explain the full range of variation found among present-day languages, particularly within closely related clusters where shared innovations and contact-induced changes have created complex patterns of similarity and difference.

Schadeberg (2009) extends this argument, contending that many linguistic developments occurred after the Proto-Bantu stage and can only be captured through reconstruction at intermediate levels. He maintains that such reconstructions are essential not only for explaining subgroup-internal developments but also for establishing more accurate genetic relationships within the family. Without intermediate reconstructions, analyses risk attributing subgroup-specific innovations directly to Proto-Bantu, thereby conflating changes from different historical periods and obscuring the actual sequence of diversification.

Campbell (2013) lends theoretical support to these arguments, contending that intermediate proto-language reconstruction constitutes a crucial but often neglected component of historical linguistics. He maintains that without intermediate reconstructions, analyses risk attributing subgroup-specific innovations directly to higher-level proto-languages, thereby distorting the historical sequence of change. Campbell's critique applies with particular force to Bantu linguistics, where the prestige and influence of Proto-Bantu reconstruction may inadvertently discourage the more localized comparative work needed to understand subgroup-level developments.

Nurse and Philippson (2003) complement this perspective by observing that intermediate reconstruction can reveal shared innovations and clarify subgroup relationships that remain opaque at broader comparative levels. They note that closely related varieties often preserve crucial evidence of historical change that becomes invisible when analysis operates solely at the Proto-Bantu level. This observation carries methodological implications: fine-grained comparison within subgroups may recover historical information that macro-level comparison necessarily loses, much as higher-resolution imaging reveals details invisible at lower resolutions.

Phonological reconstruction has occupied a central position in Bantu historical linguistics, reflecting the field's emphasis on regular sound change as the most reliable basis for historical inference. Schadeberg (2003) articulates this position clearly, arguing that systematic sound correspondences provide the foundation for reconstructing earlier language stages and for explaining patterns of diversification across the family. Through careful comparison of consonant and vowel correspondences, Schadeberg demonstrates how processes such as lenition, palatalization, vowel merger, and tonal development have operated across Bantu languages, producing the phonological diversity observable today.

Nurse (1988) makes a complementary argument focused specifically on subgroup-level analysis. Drawing on comparative data from Tanzanian Bantu languages, Nurse demonstrates that phonological comparison among closely related varieties can identify shared innovations that define genetic subgroups and reveal historical relationships not evident from broader comparison. Nevertheless, phonological studies within Bantu linguistics have predominantly remained comparative rather than explicitly reconstructive. Many analyses identify sound correspondences and describe patterns of change without taking the additional step of positing reconstructed proto-forms for specific subgroups. For language clusters such as Chagga, where phonological variation across varieties is documented, the absence of explicit reconstruction leaves historical relationships and developmental sequences inadequately explained.

2.1 Theoretical Framework

Language reconstruction relies on well-established theoretical foundations, particularly the comparative method and internal reconstruction. The comparative method, developed in the 19th century by linguists such as Franz Bopp and August Schleicher, involves comparing cognates across related languages to establish phonological correspondences and reconstruct proto-forms (Fox, 1995). A comparative method is defined by Frisco and Wilkinson (2025) as a procedure for hypothesizing a common parent language from a set of attested daughter languages. Meillet (1925) emphasizes the importance of systematic sound correspondences, arguing that linguistic changes follow regular patterns governed by phonological rules. These principles laid the groundwork for modern reconstruction techniques.

Scholars in historical linguistics generally agree that the comparative method operates through a series of systematic stages, although the number of steps varies across different theoretical presentations. For example, Campbell (2013) and Crowley (1998) present a largely five-stage model, while other scholars, such as Dimmendaal (2011), emphasize the same processes but without strict procedural numbering. Despite these differences, there is strong convergence on five core stages of reconstruction: cognate identification, establishment of sound correspondences, reconstruction of proto-forms, identification of shared innovations, and classification through tabulation of innovations. Lehmann (1992) and Fox (1995) further emphasize that reconstructed forms must be evaluated for phonological consistency, ensuring that the proposed proto-system is balanced and typologically plausible.

3. Methodology

The study adopts a qualitative approach following a widely recognized five-step methodology in historical linguistics. Data were collected from 14 informants, including two native speakers of each of the languages. The informants were selected using convenience sampling. The selection criteria included: being a native speaker of the language and being ready and available to provide the required data. The data included eight selected lexical items, which were collected through interviews. The analysis followed a systematic four-step reconstruction procedure: assembling cognate sets, establishing sound correspondences, reconstructing proto-forms, and identifying innovation. The fifth step of tabulating the innovations to determine subgroup relationships was done generally for all the words. Reconstruction was guided by two major principles: the majority principle and phonological or conventional wisdom/ phonetic plausibility.

Ethical considerations were adhered to by seeking the informants' informed consent after introducing a topic to them. Verbal consent was sought because the knowledge to read and write was not among the criteria for selection. All fourteen informants gave the data willingly.

4. The Findings

This section presents and analyzes the lexical data collected from selected Chagga varieties spoken in the Kilimanjaro region. The data consists of a list of basic vocabulary items drawn from seven varieties, namely Komochi, Kiuru, Kivunjo, Kikahe, Kiwoso, Kirombo, and Kimashami. These lexical items include core semantic domains such as kinship terms, basic actions, body parts, natural elements, and everyday nouns, which are commonly used in historical and comparative linguistic analysis. The data is organized to facilitate comparative analysis in the subsequent sections of the study. The collected words are presented in Table 1.

Table 1. Selected Lexical Items for Reconstruction

Word	Father	Dig	Hand	Die	Food	River	water	Child
Language								
<u>Komochi</u>	<i>papa</i>	<i>dema</i>	<i>owoko</i>	<i>ipfa</i>	<i>kelya</i>	<i>meda</i>	<i>mringa</i>	<i>mmana</i>
<u>Kiuru</u>	<i>papa</i>	<i>rema</i>	<i>owoko</i>	<i>ipfa</i>	<i>kela</i>	<i>mnda</i>	<i>mringa</i>	<i>mana</i>
<u>Kivunjo</u>	<i>papa</i>	<i>nlima</i>	<i>owoko</i>	<i>ifa</i>	<i>kela</i>	<i>mfoo</i>	<i>mringa</i>	<i>mana</i>
<u>Kikahe</u>	<i>papa</i>	<i>dema</i>	<i>kuoko</i>	<i>ipa</i>	<i>keja</i>	<i>dahu</i>	<i>mringa</i>	<i>mwana</i>
<u>Kiwoso</u>	<i>tata</i>	<i>dema</i>	<i>ooko</i>	<i>ipfa</i>	<i>kelya</i>	<i>ntoni</i>	<i>mda</i>	<i>muna</i>
<u>Kirombo</u>	<i>papa</i>	<i>tema</i>	<i>koko</i>	<i>ifa</i>	<i>kelya</i>	<i>kyoru</i>	<i>mryinga</i>	<i>mwana</i>
<u>Kimashami</u>	<i>baaba</i>	<i>ghema</i>	<i>oko</i>	<i>ifwa</i>	<i>shonga</i>	<i>nto</i>	<i>mugha</i>	<i>mona</i>

Table 1 presents 8 lexical items from the seven languages spoken by the Chagga people. These are presented as a representative of many other lexical items with similar structures. By reconstructing the eight lexical items, other researchers will have the basis to produce a wide range of reconstructions that will result in an ancestral language of the Chagga people, who are believed to have descended from one ancestor. The following are the step by step reconstruction of the terms.

4.1 Reconstructing the Proto Chagga Term for 'Father'

The researchers collected data for the term used to refer to father in the 7 languages and applied the reconstruction processes to all as follows.

Step One: Assemble the Cognates

The lexical items for father were collected from all seven languages. Six out of seven languages reflect cognate relations, with one having a different word. Thus, in assembling cognates, Kiwoso language was found to have a

different term, as can be seen in Table 1. Among the remaining six languages, five use the same form, with Kimashami having a slightly different term. The process continues with the six languages, leaving Kiwoso for elaboration in the latter steps.

Step 2: Create Sound Correspondences

The second step in the reconstruction of proto-terms is to create sound correspondences. In this step, sounds at the same structural position are placed in the same column before the reconstruction is done. Consonants shall be placed with consonants and vice versa. The correspondences are presented in Table 2.

Table 2. Sound Correspondences for the Terms for Father

Komochi	p	a	p	a
Kiuru	p	a	p	a
Kivunjo	p	a	p	a
Kikahe	p	a	p	a
Kirombo	p	a	p	a
Kimashami	b	a:	b	a

In this step, the terms are split into sounds, and the correspondences are presented. This step is followed by reconstruction. Frisco and Wilkinson (2025) argue that in this step, one identifies the systematic patterns of sound relationships. This step is followed by reconstruction, during which different principles can be applied.

Step 3: Reconstruct

Based on the sound correspondences identified in step 2, the proto-form for ‘father’ can be reconstructed by applying the majority principle. Most of the languages in the dataset, for example, Kimochi, Kiuru, Kivunjo, Kikahe, and Kirombo, show the form *papa*, while only Kimashami shows *baaba*, and Kiwoso has a different form, *tata*, which may represent lexical replacement rather than a cognate form.

According to the majority principle, the sounds that occur most frequently in the cognate set are likely to reflect the original proto-sounds; therefore, /p/ is preferred over /b/. This is also supported by the conventional principle, which suggests that a voiceless consonant is more likely to be original, since sound change commonly proceeds from voiceless to voiced, for example, (p → b) rather than the reverse. Therefore, considering these principles and the consistent vowel pattern across the cognates, the most plausible reconstructed proto-form for ‘father’ is **papa*.

Step 4: Establish innovations

After reconstructing the proto-form **papa* for ‘father’, the next step is to identify the innovations that occurred in the daughter languages relative to this proto-form. Most languages in the dataset retain the form *papa*, indicating retention of the proto-form without significant phonological change. However, Kimashami shows the form *baaba*, which reflects two phonological innovations. First, there is voicing, where the voiceless bilabial plosive /p/ in the proto-form has changed to the voiced bilabial plosive /b/ (p → b). Second, there is vowel lengthening, where the short vowel /a/ becomes the long vowel /a:/. These changes suggest that Kimashami has undergone independent phonological development from the proto-language.

In contrast, Kiwoso shows the form *tata*, which differs both phonologically and lexically from the reconstructed proto-form. The voiceless alveolar stop /t/ cannot be explained by a regular phonological process within the cognate set, suggesting that *tata* may represent lexical replacement, borrowing, or an innovation from another sub-branch of the language family.

Hence, while most languages preserve the proto-form **papa*, Kimashami exhibits phonological innovations, and Kiwoso appears to have a different lexical item that may indicate separate development or influence from another linguistic source.

4.2 Reconstructing the Proto Chagga Term for ‘Dig’

This applies the comparative procedures to another word in the dataset, ‘dig.’ different from the word for father, this word shows different though cognate lexical items.

Step 1: Assemble the Cognates for the word ‘dig.’

Three of the seven languages present the same term while four present cognates with varied sounds, especially at the word-initial position. Six among the seven languages share the vowels in all the words except Kivunjo, which has a different vowel in the first syllable. The cognates are presented hereunder.

Kimochi:	<i>dema</i>	Kikahe:	<i>dema</i>
Kiwoso:	<i>dema</i>	Kiuru:	<i>rema</i>
Kirombo:	<i>tema</i>	Kimashami:	<i>ghema</i>
Kivunjo:	<i>nlima</i>		

The languages above present slightly different lexical items for the word for ‘dig’ in Chagga languages. This shows that the languages have similar features which can be displayed in detail through sound correspondences in step 2.

Step 2: Create Sound Correspondences

The terms for ‘dig’ can be split into different sounds and presented as in Table 3 which arranges sounds occurring at the same structural position within the same column.

Table 3. Sound correspondences for the terms for ‘dig’

Kimochi	d	e	m	a
Kikahe	d	e	m	a
Kiwoso	d	e	m	a
Kiuru	r	e	m	a
Kirombo	t	e	m	a
Kimashami	gh	e	m	a
Kivunjo	n	l	i	a

Table 3 presents the sound correspondences of the terms in all seven languages. The initial consonant exhibits the differences in forms which however have some similarities in place of articulation, manner of articulation or voicing. These similarities make them be regarded as cognates as they reflect each other structurally. In order to come up with one proto term, the next step has to take place as follows.

Step 3: Reconstruct

Applying the majority principle, the initial consonant /d/ appears in three of the seven cognates, while /r/ occurs once, /t/ once, and /gh/ once, and /l/ occurs once. Based on frequency, the initial proto-consonant is likely to be *d, yielding a preliminary proto-form **dema*. The vowels /e/ and /a/, and the medial nasal consonant /m/, are consistent across the forms, supporting their inclusion in the reconstruction.

Then, applying the conventional wisdom principle, we consider that sound changes from voiceless to voiced are more common than the reverse, and fricatives rarely become plosives. The presence of /t/ in Kirombo and /gh/ in Kimashami suggests that the proto-sound may have been voiceless or a non-plosive fricative. Therefore, an alternative assumption for the proto-form is **tema*, where /t/ represents the original voiceless onset that later became /d/ in most daughter languages, and /gh/ in Kimashami reflects a phonetic innovation. The presence of /r/ and /l/ reflects a common sound change in Bantu in which most of the proto /d/ changed into /l/ (Nurse and Phillipson 2003).

This interpretation aligns with comparative Bantu evidence from Guthrie (1967), whose reconstruction of Proto-Bantu includes *t as a regular consonant that often surfaces as /d/ in daughter languages through voicing. This pattern can be seen in forms like **-tim-* ‘cultivate/dig’, which appears as **-tima* in some languages but as *-dima* or *-lima* in others, reflecting voicing (*t > d) and weakening (*t > l). this term can be associated also with the proto term **tema* ‘cut’ which may have undergone some semantic modification to mean cultivate. Across many Bantu languages, correspondences between *t and /d/ are widely attested, indicating that such alternations reflect regular historical developments rather than isolated changes.

What makes this argument particularly convincing is that the same tendency is clearly visible within Kimochi itself. For example, Proto-Bantu **-tapik-* ‘vomit’ appears as *daika*, **-tend-* ‘cut’ as *dumbuo*, **-tet-* ‘speak’ as *deda*, and **-tiid-* ‘run away’ as *dicha*. In each case, the original voiceless *t has shifted to a voiced /d/. The proto terms are drawn from Guthrie (1967-71) This is not a one-off change, but a consistent and systematic pattern in the language. Considering these principles together, the most plausible reconstruction for ‘dig’ is **tema*.

Step 4: Establish Innovations

Examining the reflexes of the reconstructed proto-form **tema* across the daughter languages reveals several systematic phonological developments. In Kimochi, Kikahe, and Kiwoso, the initial /t/ has undergone voicing,

producing /d/, while the rest of the word remains unchanged, showing a conservative retention of the medial /m/ and vowels /e, a/. This pattern directly reflects the regular sound change identified in Step 3, where Proto-Bantu *t commonly voices to /d/. The consistency of this change across these three varieties confirms that /d/ is an innovation rather than the original proto-sound.

In Kiuru, the initial consonant shifted to /r/, likely through a process of weakening, where the stop /t/ or /d/ softened into a liquid. This development aligns with the phonetic plausibility principle discussed in Step 3, where stops may weaken into liquids over time. The change of /d/ to /l/ or /r/ is common in Bantu. Kimashami displays /gh/ in place of /t/, which represents a fricativization of the original stop. Kirombo retains the original voiceless /t/, indicating a conservative preservation of the proto-form.

Kivunjo term *nlima*, diverges considerably from **tema* but it is likely to be maintaining the proto meaning and term for **tim*- ‘cultivate while all other languages take the word for cut. Since these languages are not spoken in isolation from other Bantu languages, Kivunjo is likely to have not accommodated the semantic change due to the neighboring Bantu languages.

Consequently, these patterns highlight a combination of voicing, fricativization, weakening, and potential lexical innovation, reflecting regular sound changes within the Chagga-related languages and isolated developments in more divergent dialects.

4.3 Reconstructing the Word for ‘Hand’

Another lexical item considered in this comparison is the word for ‘hand,’ which is constructed below:

Step 1: Assemble Cognates

The words for ‘hand’ were found to differ structurally in the seven languages. The three languages of Kimochi, Kiuru and Kivunjo have similar term *owoko*. The remaining four languages have different forms which however show some similarities. The cognates therefore are as follows.

Komochi:	<i>owoko</i>	Kiuru:	<i>owoko</i>
Kivunjo:	<i>owoko</i>	Kikahe:	<i>kuoko</i>
Kiwoso:	<i>ooko</i>	Kirombo:	<i>koko</i>
Kimashami:	<i>oko</i>		

Though they seem different, the words in all the seven languages are cognates as they show similarities in structure and meaning. It is argued by Salimili (2024), that in the first step one should decide that certain languages are related to one another. The languages spoken by the Chagga community are related. This is followed by step three.

Step 2: Create Sound Correspondences

In this step, sound correspondences are identified by putting together the words and tabulating the sounds in each position. The sound correspondences are presented in Table 4.

Table 4. Sound Correspondences for ‘hand’

Kimochi		o	w	o	k	o
Kiuru		o	w	o	k	o
Kivunjo		o	w	o	k	o
Kikahe	k	u		o	k	o
Kiwoso				o:	k	o
Kirombo	k			o	k	o
Kimashami		o	w	o	k	o

Table four presents the words for hand in the seven languages of the Chagga community. The sounds differ in structures though they are cognates. This suggest that the word for hand has undergone drastic changes in some varieties of Chagga as discussed in the fourth step.

Step 3: Reconstruct

Most languages in the dataset show the form *owoko*, while related variants occur in the other languages, such as *kuoko*, *ooko*, *koko*, and *oko*. Looking at the dataset, it is found that all the seven varieties possess the similarity in

the last three sounds which are /o/, /k/, and /o/. this suggests their belonging to the same subfamily and show their genetic relatedness. The other differences like absence or presence of some sounds are reflected here under.

Applying the majority principle, the sequence *w-o-k-o* appears most consistently across the data, indicating that these sounds reflect the original lexical root. In the historical reconstruction of Bantu nouns, prefixes are generally not reconstructed as part of roots because they belong to noun class morphology which needs a general reconstruction involving other nouns in the noun class to which the noun for hand belong. Therefore, the initial segments *ku-*, *k-* and *o-* are interpreted as noun class prefixes, while the remaining segment represents the lexical root.

Some forms show regular phonological changes. The absence of /w/ in forms such as *oko* and *ooko* can be explained by glide deletion ($w \rightarrow \emptyset$), a phonetically plausible process since glides are weak segments. Also, the form *kuoko* likely reflects the presence of noun class prefix *k-*, and weakening of /w/ into /u/ while *koko* in shows glide /w/ deletion.

Considering these correspondences and the dominance of the sequence *woko* in the cognate set, the most plausible proto-root for ‘hand’ is therefore reconstructed as **-woko*.

Step 4: Establish Innovations

After reconstructing the proto-root **-woko* for ‘hand’, the next step is to examine the innovations that occurred in the daughter languages relative to this root. Some languages retain the form *owoko*, which closely reflects the proto-root **-woko*, together with a noun-class prefix *o-*. These forms, therefore, represent conservative developments that preserve the original phonological structure.

Kimashami shows the form *oko*, which reflects glide deletion, where the consonant /w/ is lost ($w \rightarrow \emptyset$), a phonological process affecting weak glide sounds. Kiwoso presents *ooko*, which can be explained by vowel lengthening after the deletion of /w/, yielding the long vowel /o:/.

In Kikahe, the form *ku-oko* adds the noun class 15 prefix (*ku-*) at the beginning of the word, representing a morphological structure of the noun-class prefix. Kirombo displays the form *koko*, which reflects glide deletion ($w \rightarrow \emptyset$) from the root which also happens in Kikahe. Also, the vowel u of the noun class 15 prefix *ku-* is deleted resulting into *k-*.

So, these developments demonstrate several phonological processes, including glide deletion, vowel lengthening, prefix modification. Despite these innovations, most forms still preserve the core structure of the reconstructed proto-root **-woko*, confirming their historical relationship, while the modified forms represent later developments within particular languages.

4.4 Reconstruction of the Word for ‘Die’

The researchers collected the lexical items for the word ‘die’ from all the seven languages. The words presented a high level of similarities making them all cognates. The reconstruction of the word followed all the steps as presented below.

Step 1: Assemble the Cognates

All the seven languages represent words which possess structural and semantic similarities leading into all being assembled as cognates.

Kimochi:	<i>ipfa</i>	Kiuru:	<i>ipfa</i>
Kivunjo:	<i>ifa</i>	Kikahe:	<i>ipa</i>
Kiwoso:	<i>ipfa</i>	Kirombo:	<i>ifa</i>
Kimashami:	<i>ifwa</i>		

Step 2: Create Sound Correspondences

The words for ‘die’ were found to have unique sounds like /pf/ is three languages. The sound correspondences are presented in Table 5.

Table 5. Sound Correspondences for ‘die’

Kimochi	i	pf	a
Kiuru	i	pf	a
Kivunjo	i	f	a
Kikahe	i	p	a
Kiwoso	i	pf	a
Kirombo	i	f	a
Kimashami	i	f ^w	a

Step 3: Reconstruct

The reconstruction of the proto-form is guided by the majority principle and conventional (phonological) wisdom/phonetic plausibility. Most languages in the dataset: Kimochi, Kiuru, and Kiwoso, show the form *ipfa*, while related variants appear in the other languages, such as *ifa* in Kivunjo and Kirombo, *ipa* in Kikahe, and *ifwa* in Kimashami.

Applying the majority principle, the voiceless labio-dental affricate /pf/ occurs most frequently in the medial position, suggesting that this sound reflect the original proto-structure. The forms *ifa* in Kivunjo and Kirombo are less frequent while /p/ and /f^w/ appear only once each. Thus, **ipfa* is suggested to be a proto form by majority principle. Taking into account conventional wisdom, the plosive /p/ is stronger than /pf/ making /p/ the possible proto sound. Thus by conventional wisdom, the form *ipa* in Kikahe is proto. Considering the consistent vowel pattern /i/ and /a/ across the cognates and the phonetic strength of the voiceless bilabial plosive /p/ the most plausible reconstructed proto-form is **ipa*.

Step 4: Establish Innovations

In analyzing the cognates for the word ‘die’ across the Chagga languages, several phonological innovations relative to the reconstructed proto-form **ipa* can be observed. Kivunjo retain the proto form **ipa* while the other varieties diverge slightly from the same. Kimochi, Kiuru, and Kiwoso reflect affrication in which the voiceless bilabial plosive /p/ changes into a voiceless labio-dental affricate /pf/ with Kirombo reflecting fricativization of the same plosive into a voiceless labio-dental fricative /f/. Kimashami exhibits a change from /p/ to /f^w/ in *ifwa*, which represents a combination of glide formation /w/ and fricativization (/f/ remains), suggesting a phonological innovation that differentiates it from the rest of the languages.

4.5 Reconstruction of the Word for ‘Food’

Another lexical item from the data is ‘food.’ This is also analyzed as follows;

Step 1: Assemble the Cognates for the word ‘food’

The cognates of the word for ‘eat’ in the seven languages are assembled as follows

Kimochi:	<i>kelya</i>	Kiuru	<i>kela</i>
Kivunjo:	<i>kela</i>	Kikahe	<i>keja</i>
Kiwoso:	<i>kelya</i>	Kirombo	<i>kelya</i>

The data shows that six among the seven languages share cognates while Kimashami uses a completely different term for food (*shonga*). This will be discussed in step four of the reconstruction. The words in the six languages are not completely similar, they differ in some ways as discussed in this section.

Step 2: Create Sound Correspondences

The sound correspondences for the cognates in the six languages are presented in Table 6.

Table 6. Sound correspondences for the word ‘eat’

Kimochi	k	e	l	y	a
Kiuru	k	e	l		a
Kivunjo	k	e	l		a
Kikahe	k	e	j		a
Kiwoso	k	e	l	y	a
Kirombo	k	e	l	y	a

The languages show great similarities of sounds in four of five structural positions. This evidences that they descend from the same ancestor and that their immediate proto language is one.

Step 3: Reconstruct

The reconstruction of the proto-form for ‘food’ is also guided by the majority principle, conventional wisdom, and phonetic plausibility. Applying the majority principle, the structure *k-e-l-y-a* appears in several languages, which show the form *kelya*. Kiuru and Kivunjo show *kela*, while Kikahe shows *keja*. Because the consonants /k/, /e/, /l/, and vowel /a/ occur consistently across most forms, and the glide /y/ appears in several languages, the first assumption is that the proto-form was **kelya*.

According to conventional wisdom, the correspondence between /l/ and the palatal glide /j/ suggests that /l/ is more likely the original sound, since lateral consonants may weaken or shift toward palatal glides through palatalization. Thus, the development /l→j/ explains the Kikahe form *keja*. In addition, the absence of /y/ in the forms *kela* can be explained by glide deletion ($y \rightarrow \emptyset$), which is a common phonological process. These patterns support the reconstruction **kelya*.

Step 4: Establish Innovations

In this, Kimochi, Kiwoso, and Kirombo retain the form *kelya*, showing no significant phonological change and therefore representing conservative preservation of the proto-form. Kiuru and Kivunjo show the form *kela*, where the glide /y/ has been deleted ($y \rightarrow \emptyset$). This reflects glide deletion, a common phonological simplification in which a weak glide is lost while the remaining structure of the word is preserved.

Kikahe presents the form *keja*, where the lateral /l/ changes to the palatal consonant /j/ ($l \rightarrow j$). This development represents palatalization, a phonological innovation that alters the place of articulation of the consonant while maintaining the overall structure of the word.

The form *shonga* in Kimashami differs considerably from the other forms. Its consonant and vowel structure does not correspond to the reconstructed proto-form *kelya*. Because it does not follow the regular sound correspondences observed in the other languages, it may represent lexical replacement, borrowing, or an independent innovation within this language.

So, the innovations identified in this dataset include glide deletion (*kelya* → *kela*) and palatalization ($l \rightarrow j$), while the form *shonga* represents a lexical divergence from the reconstructed proto-form. These developments illustrate how the daughter languages have diverged phonologically from the proto-language while still maintaining identifiable relationships among most of the forms.

4.6 Reconstruction of the Term for ‘River’

Having completed the reconstruction of the word ‘food’, the analysis now proceeds to the next lexical item, ‘river.’ River was found to be represented by different lexical items some of which reflecting Kiswahili, a national language form and other unique to specific languages. Thus, the assembling of cognates was complex with the complex process of finding the Proto-Chagga term for river.

Step 1: Assemble the Cognates for the Word for ‘River’

The data are divided into four groups with Kimochi and Kiuru having structurally similar forms. Kivunjo, Kimashami and Kiwoso also have similar forms with Kikahe and Kirombo having unique form each. The differences may result from the historical background of the Chagga people who found the river to be sacred or the other cultural significance associated with rivers in the community.

Kimochi:	<i>meda</i>	Kivunjo:	<i>mfoo</i>
Kiuru:	<i>mnda</i>	Kiwoso:	<i>ntoni</i>
Kikahe:	<i>dahu</i>	Kimashami:	<i>nto</i>
Kirombo:	<i>kyoru</i>		

The dataset above presents the words in the seven languages. There are less cognates in these languages which show the possibility of lacking a word for river in the proto Chagga language. It is argued by Kim, et al (2023) that historical linguistics reconstruct proto- languages by identifying systematic sound changes that can be inferred from correspondences between attested daughter languages. Thus, having no cognates for the word for river makes it complex to reconstruct a single term for ‘river’ in these languages.

The languages Kimochi, and Kiuru forms show similarities. Kimashami uses the form *nto*, while Kiwoso uses *ntoni*, in which *nto* receives the suffix *-ni* to form *ntoni*. This represents suffixation or morphological extension, while the core sequence /nto/ remains intact. Kivunjo term *mfoo* can be compared to *nto* due to the structural and semantic similarity with both having the nasal onsets. The labia dental fricative /f/ can be said to be formed through frication of the voiceless alveolar stop /t/ while /o:/ can be said to result from the lengthening of /o/. The

shared structure indicates a close relationship between Kiwoso, Kimashami and Kivunjo, forming another subgroup within the dataset. The forms *dahu* in Kikahe and *kyoru* in Kirombo do not follow the sound correspondences observed in the other sets. Their consonant and vowel patterns differ significantly from each of the two groups above.

Thus the term for ‘river’ cannot be traced in Chagga Uru through sound correspondences. Historical traces may be needed to reach the conclusion about the significant disparities of the words for river in these languages despite being sisters. This can be done also by adding other Bantu languages to see the trend. River is among the significantly divine places where rituals were previously conducted in different Chagga communities. Generally, the presence of multiple lexical forms for ‘river’ across these languages cannot be explained solely by phonological differences; cultural and semantic factors also play a role. This calls for a detailed study of names of places at which rituals are conducted among the Chagga communities to see the trend.

In many Bantu-speaking communities, natural features such as rivers are not merely physical elements of the landscape; they carry important social, ecological, and spiritual meanings. Rivers are often linked to ritual practices, beliefs about ancestors, and sacred spaces, and these associations can influence how people name and talk about them. As studies of African worldviews, such as *African Religions and Philosophy* by Mbiti (1969), show, elements of the natural environment frequently have religious and symbolic value, which can lead to variation in linguistic expression. In addition, research on Bantu languages shows that culturally important areas of life often develop a richer, more specialized vocabulary (Nurse & Philippson, 2003). For this reason, the variation seen in the word ‘river’ is likely not just the result of historical sound changes but also reflects cultural differences, including the use of different terms for different types of rivers or for specific social and cultural contexts.

4.7 Reconstructing the Word for ‘Water’

The data set in Table 1 includes the word for water in the seven languages. The word was found from all seven languages. The cognates are assembled as follows.

Step 1: Assemble the Cognates for the word ‘Water’

The words for water were found to consistent in five of the seven Chagga languages. In Kimashami and Kiwoso, the words are similar in some ways and different in others. The words in the two languages share some similarities including the initial nasal sound and the final vowel. Thus they are taken as cognates as follows.

Kimochi:	<i>mringa</i>	Kiuru:	<i>mringa</i>
Kivunjo:	<i>mringa</i>	Kikahe:	<i>mringa</i>
Kirombo:	<i>mryinga</i>	Kimashami:	<i>mugha</i>
Kiwoso:	<i>mda</i>		

The consistence of the words in the above dataset reflects the genetic relationship among the languages while their differences shows their departure from the same. This step is also applied by Ferdnand (2024) when reconstructing the proto terms of three sister languages of Luganda, Ruhaya and Kiswahili.

Step 2: Create Sound Correspondences

The sounds making up words for ‘water’ are clearly identifiable. They include the noun class prefix which seem to be uniform throughout the languages. It is argued by Ferdnand (2024) that, while Oruhaya and Oluganga nouns begin with vowels (augments) Kiswahili nouns do not show such structure as they start with the noun class prefix. This shows that the morphological structure of words may differ from language to language even when they are sisters. Sound correspondences for the words are presented here under.

Table 7. Sound Correspondences for ‘water’

Kimochi	m	r	i	ng	a
Kiuru	m	r	i	ng	a
Kivunjo	m	r	i	ng	a
Kikahe	m	r	i	ng	a
Kirombo	m	ry	i	ng	a
Kimashami	m		u	gh	a
Kiwoso	m			d	a

Table 7 presents the sound correspondences for the words in all seven languages. The words are nouns and start with the noun class three prefix *m-*. The languages share the noun class prefix and non have the augment. The noun class prefixes are followed by the roots which are completely similar in Kimochi, Kiuru, Kimochi, and Kikahe. Kirombo term *mryinga* presents a slightly different root while the roots Kimashami and Kiwoso roots are highly different.

Step 3: Reconstruct

The reconstruction takes into account the sounds in step 2. Applying the majority principle, the sequence *m-r-i-ng-a* appears four times. This consistent correspondence suggests that these sounds reflect the earlier structure of the word. Based on this distribution, the first assumption is that the proto-form was ***mringa**.

Following conventional wisdom, simpler consonant structures are generally considered original, while additional segments often arise through later phonological developments. The form *mryinga* shows the sequence /ry/, which can be explained as palatalization or glide insertion from an earlier /r/. The form *mugha* differs through changes in both vowel and consonant structure, where /r/ is lost and the cluster /ng/ corresponds to /gh/. This suggests fricativization and vowel modification, indicating later innovation rather than the original form.

The phonetic plausibility principle also helps explain the remaining variations. Kiwoso shows *mda*, which can be interpreted as the result of the deletion of the medial segments /r, i / followed by the change from /ng/ to /d/. Since the majority of languages preserve the sequence *mringa*, and the other forms can be explained through different processes, the most plausible proto-root for ‘water’ is therefore reconstructed as ***-ringa**, while *m-* is interpreted as a noun class 3 prefix rather than part of the lexical root.

Step 4: Establish the Innovations

After reconstructing the proto-form ***-ringa**, we can identify the innovations that occurred in the daughter languages relative to this proto-form. Several phonological changes are obvious. Kimochi, Kiuru, Kivunjo, and Kikahe retain the structure *mringa*, indicating that these languages preserve the proto-form with little or no phonological modification. In Kiwoso, the form *mda* reflects a significant reduction. The segments /ri/ appear to have been deleted and /ng/ changed to /d/, producing a simplified structure *mda*. This development suggests consonant and vowel deletion accompanied by cluster simplification, where the complex sequence found in the proto-form is reduced to a shorter pattern.

In Kirombo, the word *mryinga* shows glide insertion or palatalization, where the consonant /r/ develops into /ry/. This represents a phonological weakening or palatal development while the remaining segments /i, ng, a/ remain stable, indicating partial retention of the earlier structure. Kimashami, on the other hand, displays the form *mugha*, which diverges more strongly from the proto-form. Here, the vowel /i/ changes to /u/, representing vowel modification, while the nasal cluster /ng/ develops into the fricative /gh/, suggesting fricativization and consonant substitution. The loss of /r/ further reflects segment deletion.

Although most languages clearly reflect the proto-root ***-ringa**, the form in Kimashami is structurally distant from the other cognates. Such divergence may point to lexical innovation, borrowing from another language, or independent phonological development within that branch of the family.

4.8 Reconstructing the Word for ‘Child’

The researchers collected the data for the word for ‘child’ from all the seven languages. The comparative method is applied to the word child following the steps below.

Step 1: Assemble the Cognates

The words in all seven languages present structural and semantic similarities with slight differences. The cognates are presented here under

Kimochi:	<i>mmana</i>	Kiuru:	<i>mana</i>
Kivunjo:	<i>mana</i>	Kikahe:	<i>mwana</i>
Kiwoso:	<i>muna</i>	Kirombo:	<i>mwana</i>
Kimashami:	<i>mona</i>		

The cognates in the sister languages above show some differences which may have been the result of phonological changes. This is because they are still preserving some common features. It is argued by Munteanu (2022) that reconstruction can be done using a set of synchronically equivalent word lists which refer to the words which share the same phonological and phonotactic properties. They are the words which share the segmental inventory of the daughter. The sister languages above are synchronic and they share some phonological properties. The sound correspondences are as follows.

Step 2: Create Sound Correspondences

The sounds making up the word for child are arranged depending on the position they occupy. This arrangement however takes into consideration the noun class prefix, in this case noun class one. The reconstructed proto Bantu noun class one prefix is **mu-* (Guthrie 1967-1971: 144) and the root for child is **-na* or **-yana* (p. 148). Thus, the arrangement of the sound follows this schema since all the seven languages are Bantu. Table 8 presents the sound correspondences.

Table 8. Sound Correspondences for the word for ‘child’

Kimochi	m	m	a	n	a	
Kiuru		m	a	n	a	
Kivunjo		m	a	n	a	
Kikahe		m	u	a	n	a
Kiwozo		m	u		n	a
Kirombo		m	u	a	n	a
Kimashami		m	o		n	a

Table 8 presents the sound correspondences with each sound on its structural position. The words in all the seven languages start with the bilabial nasal m- with Kimochi having two nasals at the same position. The nasals are followed by the glide /w/ in Kikahe and Kiwozo and /u/ and /o/ in Kirombo and Kimashami respectively. The vowel /a/ is not placed in such position due to the presence of the same in other languages making it to occupy a separate slot. The final two sounds /n/ and /a/ are consistently present in all the languages.

Step 3: Reconstruct

Applying the majority principle, the sequence *-ana* / *-na* appears most consistently across the languages and represents the stable lexical core, while the initial consonant *m-* or *mu-* is treated as a noun class prefix which reflects Guthrie’s reconstructed noun class one prefix. Thus, considering these correspondences the most plausible proto word for ‘child’ is **mu-ana*

Step 4: Establish the Innovations

Based on the reconstructed proto-root **mu-ana* for ‘child’, several phonological innovations are observable in the daughter languages. Kikahe and Kirombo retain the proto-root directly as *mwana* (*mu-ana*) with the glide formed from the combination of /u/ and /a/. Gliding is argued by Mpobela (2025) to be a common morphophonological process in Bantu which is used to solve hiatus. She presents the same pattern in Runyambo, a Bantu language spoken in north western Tanzania. Kiuru and Kivunjo present *m-ana* in which the vowel /u/ is deleted from the noun class one prefix. Kimochi shows *mmana*, where the initial consonant /m/ is geminated, reflecting consonant strengthening.

Kiwozo has *muna*, in which the first vowel of the root is deleted or the language takes the root from the reconstructed Bantu root *-na*. Similarly, Kimashami presents the form *mona* (analyzable as *mu-* + *ana*), where the vowels /u/ and /a/ coalesce to form /o/. Sibanda (2008) presents coalescence as one of the solutions to **vv* in Nguni. Coalescence is defined by Sibanda as the process in which two different vowels form a single monomoraic vowel with non-conflicting features from the two vowels that combine. Thus in Kimochi the two vowels /u/ and /a/ combine to form /o/.

A closer examination of the data reveals that the noun class 1 prefix is not realized uniformly across the languages but appears in several phonological variants. These include *m-* (Kiuru, Kivunjo), *mm-* (Kimochi), and *mu-* (Kikahe, Kirombo, Kiwozo, Kimashami). These forms can be understood as allomorphs of the same underlying noun class prefix, rather than distinct morphemes. The variation is conditioned by phonological environments, particularly the quality of the following vowel, as well as language-specific phonological developments. For example, the form *mw-* before the vowel /a/ can be analyzed as glide formation, while *mu-* and *mo-* reflect vowel modification within the prefix and the root. Such variation is widely attested in Bantu languages, where noun class prefixes show systematic phonological alternations across related languages (Guthrie, 1971). This supports the analysis that the prefix belongs to noun class 1 and should be treated as a grammatical element rather than part of the lexical root.

Taken together, the innovations relative to **mu-ana* include consonant germination, glide formation, vowel deletion and vowel coalescence. All forms retain the core consonant framework *-n-a*, so none diverge drastically

from the proto-root, confirming that they belong to the same ancestral language rather than representing borrowings or unrelated words.

4.9 Tabulate the Innovations for all the Reconstructed Terms

Based on the innovations identified in step four, the data shows that the seven languages belong to the same ancestral mother language (Proto-Chagga). While each of the seven languages retain the proto form in one word or the other, Kiwoso, Kivunjo and Kimashami present some words which seem not to belong to the Proto Chagga language. Kiwoso shows the form *tata* ‘father’ which differs significantly from the reconstructed Proto-Chagga form (*papa) both phonologically and lexically. The change from /p/ to /t/ does not follow the regular sound correspondences observed in the other languages, suggesting that *tata* may represent lexical replacement, borrowing, or influence from another sub-family.

For the word *tema* ‘cultivate’, Kimochi, Kikahe, and Kiwoso share the same voicing of /t/ to /d/ and maintain the medial /m/ and vowels /e,a/, forming a conservative cluster within the family. Kirombo, which retains the original /t/ constitutes a separate branch preserving the proto-form, indicating minimal phonological change. Its retention of /t/ supports the reconstruction of *tema, since conservative forms are typically closer to the proto-language. The conventional wisdom also suggests the stronger sound to be proto, making it commendable for /t/ to be proto rather than /d/.

For the word for ‘hand’, the reconstructed proto-root is *-woko. Kimochi, Kiuru, and Kivunjo form preserve the form -woko, which clearly reflects the proto-root together with the noun class prefix o-. Kimashami presents the form oko, which differs slightly through glide deletion (w → Ø). Kiwoso term ooko, present glide deletion followed by vowel lengthening, resulting in the long vowel /o:/. This development represents a distinct phonological innovation affecting the word's vowel structure. The forms kuoko and koko in Kikahe and Kirombo respectively reflect the presence of a noun class 15 prefix ku- appearing at the beginning of a noun. On the same prefix, the high vowel of the prefix -ku is deleted in Kirombo. This change reflects variation in the noun class prefix rather than the lexical root itself. The data in the two languages also show glide deletion.

The reconstructed form die is *ipa which is found in only one language which is Kikahe. From the reconstructed proto-form *ipa, the cognates for the word ‘die’ show clear patterns of phonological innovation that allow us to establish an internal classification of the languages. Kimochi, Kiuru, and Kiwoso modify the term to ipfa. Kivunjo and Kirombo exhibit the fricativisation of the proto */p/ to /f/ weakening, which sets them apart as a distinct subgroup. Kimashami demonstrates a more complex change including fricativization and gliding into if^wa, (p → f^w). These patterns highlight sound changes, mainly affrication, fricativization and glide formation, that can be used to construct an internal family tree reflecting the phylogenetic relationships of the languages.

The reconstructed proto-form *kelya ‘eat’ reveals patterns that help establish subgroup relationships among the languages. Kimochi, Kiwoso, and Kirombo retain the form kelya, showing no observable phonological change. Kiuru and Kivunjo show the form kela, where the glide /y/ has been deleted (y → Ø) while Kikahe shows the form keja, in which hardening and palatalization has occurred resulting into a palatal affricate /dʒ/. The form shonga in Kimashami differs significantly from the forms in all other languages which may have resulted from borrowing or generalization of the form that was previously used for a specific delicacy.

For the term for ‘river’, we can generally say that some terms presents varied forms making it difficult to tabulate innovations based on a single word. While all the other words suggests the Chagga languages genetically belonging to one language group, the term for ‘river’ if used in isolation may suggest differently. This suggests that, some terms have diverged less from the reconstructed proto term while others have been totally replaced by other terms.

For the proto-form *-ringa ‘water’, Kimochi, Kiuru, Kivunjo, and Kikahe retain the proto term mringa, without any phonological modification. Kiwoso, presents a reduced form the mda in which the syllable /ri/ of the reconstructed proto term is deleted and /ng/ is represented by /d/. The word mryinga Kirombo shows the phonological change of palatalization/gliding of /r/ into /ry/ while Kimashami form mugha diverges the proto-form. Thus, while other languages maintain the proto terms or the cognates, Kimashami and Kiwoso have changed drastically from the reconstructed proto term. This suggests further research into the neighbouring Bantu and non languages.

For the word for ‘child’, Kikahe and Kirombo share the formation of the glide /w/ between by combining /u/ and /a/, resulting in mwana, which represents gliding or labialization and forms a gliding subgroup. Kiuru and Kivunjo delete the final vowel /u/ of the noun class one prefix resulting into mana. Kimochi shows mmana, where the initial nasal /m/ is geminated. Kiwoso has muna, with the first vowel /a/ of the root deleted, while Kimashami shows mona, with the vowels /u/ and /a/ coalescing into /o/.

Generally, based on the comparative analysis of the Chagga cognates, the internal classification of the languages can be explained through shared retention and innovations, guided by specific assumptions. All the seven

languages belong to the same ancestral mother language, in this case Proto- Chagga. A few differences exist possibly as a result of contact with neighbouring Bantu and non Bantu languages but most of the languages preserve the cognates which relates mostly to the reconstructed proto terms. All the words have reconstructed proto terms except a word for river which is represented by diverse terminologies in the seven languages.

5. Conclusion

Generally, the Chagga people seem to have had one language before it was split into the present seven varieties. This is reached due to the similarities discussed in this paper. Despite the differences, there is still a lot of similarities in structure and meaning which however may need a more comprehensive study in which more data will be tested and validated. This is similar to the argument by Mabugu, et al (2024) on Shona which they there was one variety of the same before it split into a variety of dialects. They believe that Kalanga was the original proto language of Shona.

The analysis revealed that many of the Chagga varieties share regular sound correspondences, confirming their historical relationship. For example, the proto-form **papa* ‘father’ was retained in most varieties, while Kimashami showed innovations such as voicing and vowel lengthening, and Kiwoso exhibited lexical replacement. Similarly, the reconstruction of **tema* ‘dig’, **-woko* ‘hand’, **ipa* ‘die’, **kelya* ‘food’, and **-ringa* ‘water’ demonstrated consistent phonological patterns across the languages.

Furthermore, the findings indicate that several phonological processes have contributed to the diversification of the Chagga varieties. These include glide deletion, vowel deletion, vowel lengthening, palatalization, consonant weakening, fricativization, and lexical replacement. These processes explain the variations observed in the daughter languages while still reflecting their common origin.

References

- Campbell, L. (2013). *Historical linguistics: An introduction* (3rd ed.). Edinburgh University Press.
- Ehret, C. (2001). Bantu expansions: Re-envisioning a central problem of early African history. *The International Journal of African Historical Studies*, 34(1), 5-41.
- Ferdinand, A. (2024). A Comparative Reconstruction of Sister languages: Oruhaya, Oluganda and Kiswahili. *Journal of Linguistics and Language in Education*. 18(1), 123-131. <https://doi.org/10.56279/jlle.b18i1.8>
- Fox, A. (1995). *Linguistic reconstruction: An introduction to theory and method*. Oxford University Press.
- Frisco, A., & Wilkinson, C. (2025). Reconstructing the unwritten: Methodological advances and challenges in the study of ancient languages. *Language Perspectives*, 1(1), 1-7. <https://doi.org/10.644229/chg34d13>
- Guthrie, M. (1967). *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages* (Vols. 1-4). Gregg International.
- Guthrie, M. (1967-1971) *Comparative Bantu: An introduction to the comparative Linguistics and prehistory of the Bantu languages*, 2: *Bantu prehistory, inventory and indexes*. London. Gregg International.
- Hock, H. H. (1991). *Principles of historical linguistics* (2nd ed.). Mouton de Gruyter.
- Hock, H. H., & Joseph, B. D. (2009). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (2nd ed.). Mouton de Gruyter.
- Hock, H. H., & Joseph, B. D. (2019). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (3rd ed.). De Gruyter Mouton.
- Kim, Y. M., Chang, K., Cui, C., & Mortensen, D. (2023). Transformed Protoform Reconstruction. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics held from 9th to 14 July*. 2, 24-38.
- Mabugu, P. R., Mutonga, L., & Mabvumo, M. (2024). Reconstruction of Proto-Shona: A Case of Karanga. *Journal of African Languages and Literary Studies*, 5(3), 25-40. <https://doi.org/10.31920/2633-2116/2024/v5n3a2>
- Mbiti, J. S. (1969). *African religions and philosophy*. London: Heinemann.
- Meeussen, A. E. (1967). Bantu grammatical reconstructions. *Africana Linguistica*, 3, 79-121.
- Mpobela, L. (2022). 7 to 5 Vowel Shift: Scars of the Lost Vowels in Runyambo. *International Journal of Literature and Arts*, 10(5), 278-291 <https://doi.org/10.11648/j.ijla.20221005.13>

- Mpobela, L. (2025). Sound Patterns in the Compound Words in Runyamigongo Dialect of Runyambo. *Linguistic Exploration*, 2(2). <https://doi.org/10.55121/le.v2i2.724>
- Munteanu, A. (2022). Comparative Reconstruction Probablistically: The role of Inventory and Phonotactics. *Proceedings of AMP 2021*, 1-11. <https://doi.org/10.3765/amp.v9i0.5168>
- Nurse, D. (1988). The diachronic background to the language communities of southwestern Tanzania. *Sprache und Geschichte in Afrika*, 9, 15-55.
- Nurse, D., & Philippson, G. (Eds.). (2003). *The Bantu languages*. Routledge.
- Philippson, G. (2006). Bantu languages. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 1, pp. 670-676). Elsevier.
- Salimili, E. M. (2024). The Comparative Method in the Reconstruction of Proto-Languages. *Western Caspian University, Scientific Bulletin*, 2, 10-18.
- Schadeberg, T. C. (2003). Historical linguistics. In D. Nurse, & G. Philippson (Eds.), *The Bantu languages* (pp. 143-163). Routledge.
- Schadeberg, T. C. (2009). Bantu languages. In B. Comrie (Ed.), *The world's major languages* (2nd ed., pp. 609-627). Routledge.
- Sibanda, G. (2009). Vowel Processes in Nguni: Resolving the Problem of Unacceptable VV Sequences. In Masangu, M., Laughlin, F.C., & Potsdam, E. (Eds.), *Selected Proceedings of the 38th Annual Conference on African Linguistics. Cascadilla Proceedings Project* (pp. 38-55). Somerville, MA, USA.
- Trask, R. L. (2000). *The dictionary of historical and comparative linguistics*. Fitzroy Dearborn Publishers.
- Trask, R. L. (2015). *Trask's historical linguistics* (3rd ed., revised by R. McColl Millar). Routledge.

Copyrights

The journal retains exclusive first publication rights to this original, unpublished manuscript, which remains the authors' intellectual property. As an open-access journal, it permits non-commercial sharing with attribution under the Creative Commons Attribution 4.0 International License (CC BY 4.0), complying with COPE (Committee on Publication Ethics) guidelines. All content is archived in public repositories to ensure transparency and accessibility.