

## Linguistic background to the Mara Bantu corpus languages

Tim Roth

The Mara Region of Tanzania is home to an incredibly diverse and complex linguistic landscape. At least twenty Bantu language varieties are spoken in the region, along with an additional three Nilotic languages (Datooga, Luo, and Maasai) with multiple varieties of their own. The Mara Region is located in the northwestern part of Tanzania with Lake Victoria to the west, Serengeti National Park to the south and east, and Kenya to the north.

The Mara Bantu languages in the Shetler corpus include Ikizu, Ikoma, Ishenyi, Jita, Kuria, Nata, Ngoreme, Zanaki, and Temi. (While Temi is actually spoken in the geographically adjacent Arusha Region, it has close linguistic and socio-historical connections to the Western Serengeti languages in the Mara Region). The Bantu languages are considered part of the much larger *Niger-Congo* language family. Niger-Congo languages are spoken across the majority of sub-Saharan Africa, with the Bantu languages forming one of the larger branches of the Niger-Congo family tree. Estimates of the number of Bantu languages vary because experts often have different criteria regarding what constitutes a *language* versus a *dialect*. However, there are around 500 different Bantu languages from Nigeria/Cameroon in the west, to Kenya/Tanzania in the east, and all the way to South Africa. Some of the more well-known Bantu languages include Swahili, Chichewa, Gikuyu, Kinyarwanda, Luganda, Xhosa, and Zulu. Linguists who specialize in Bantu languages often use a geographical classification system consisting of a capital letter (*zone*) followed by a two- or three-digit number called a *Guthrie code* to aid in identifying Bantu languages, e.g. Temi (E.46). Around Lake Victoria, a subsequent update added a zone J to some languages within existing zones D and E to highlight potential close genetic connections, e.g. Kuria (JE.43).

The Bantu languages in the Mara region are part of a branch of Bantu languages called *Great Lakes* Bantu. In the diagram below, we can see that the Great Lakes branch is split into four different subgroups: *Western Lakes*, *West Nyanza*, *East Nyanza*, and *Greater Luhyia*.

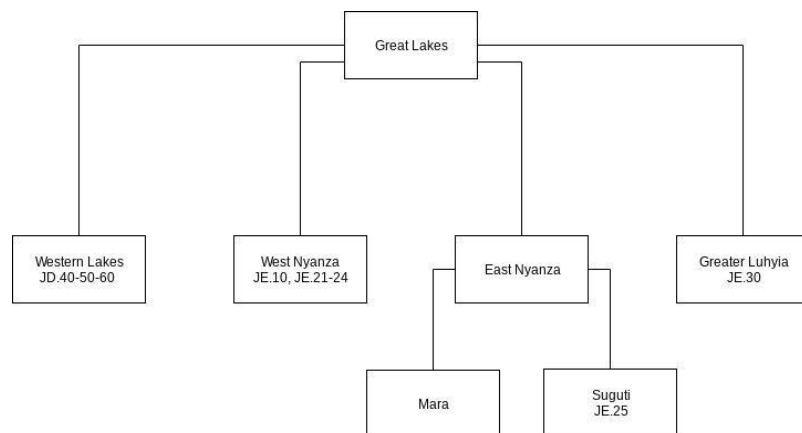


Figure 1. Great Lakes family tree diagram (see Schoenbrun 1997: 12-13)

What we have so far been referring to as the Mara Bantu languages fall under East Nyanza – in the diagram both *Mara* and *Suguti* languages. (The sole Suguti language in the Shetler corpus is Jita).

Linguists typically further divide this group of Mara languages into two subgroups: *North Mara* and *South Mara*. The remainder of the Mara Bantu languages within the Shetler corpus (with the exception of Kuria) are South Mara languages. See Figure 2 below.

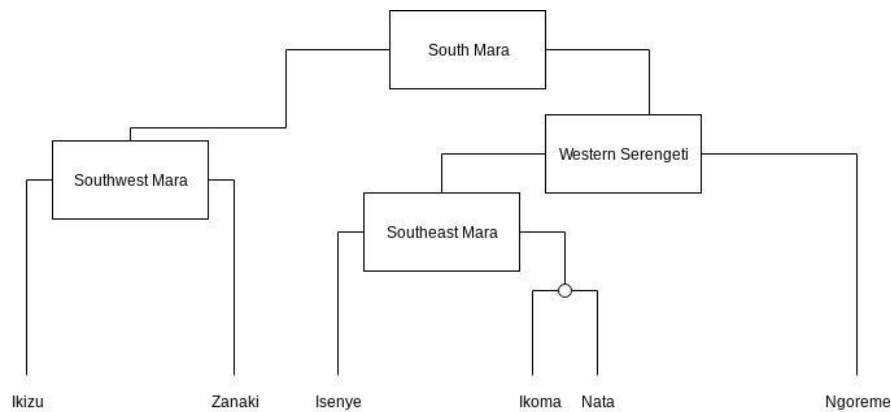


Figure 2. South Mara tree diagram (Roth 2018)

These groupings are not fixed/absolute, and often change based on new data or insights by linguists and historians. Some linguists and historians arrive at these groupings based solely on the vocabulary (or *lexicon*) of a given set of languages, while others incorporate data from other parts of the language in addition to the lexicon, including *phonology* (sounds), *morphology* (words), and *syntax* (sentences). Groupings based only on the lexicon can often be in error because of languages borrowing words from each other (or more broadly *language contact*). Borrowing can disguise the degree of close relationship between languages. However, this doesn't mean the lexicon isn't important, it just means that this can be a very complicated process for linguists and historians to work out.

For example, in the East African context Swahili is a language of wider communication. Swahili (G.42) is also a Bantu language, but is part of a different branch on the tree than the Mara Bantu languages. For example, let's look at Table 1 below which lists the words for *nyoka* 'snake' in the Mara corpus languages, the singular forms on the left and the plural on the right.

Table 1. Mara Bantu reflexes for \*-jóká 'snake'

<b>Jita</b>	indʒoka / dʒindʒoka
<b>Kuria</b>	inchoka / ichinchoka
<b>Ngoreme</b>	enchoka / chinchoka
<b>Ikoma</b>	anchoka / chanchoka
<b>Isenye</b>	inchoka / chinchoka
<b>Nata</b>	anchoka / chanchoka
<b>Ikizu</b>	enzoka / zenzoka
<b>Zanaki</b>	enzoka / ezinzoka

From just an initial observation of the Mara words, they each have the *-oka* portion that is also present in the Swahili word *nyoka*. Is this because the Mara languages borrowed this word from Swahili, because Swahili and these Mara languages are all Bantu languages, or is it just coincidence? In this particular example, the Proto-Bantu word for ‘snake’ is *\*-jóká*, and this is the reason why both Swahili and the Mara Bantu languages have such similar wordforms. In turn, sometimes the similarities are just coincidence, and sometimes the Swahili word has been borrowed into many different languages.

This *nyoka* example doesn’t tell us much about the internal relationships between the Mara languages however. (Although perhaps we could notice that all but Jita, Ikizu, and Zanaki have the *-choka* root). In Table 2 below, there is a list of the words for *mbuzi* ‘goat’ in the Mara corpus languages. In this example, the origin of the word again is Proto-Bantu, in this case *\*-bódi*. For reference’s sake, the word for ‘goat’ in Temi is *mbori*.

Table 2. Mara Bantu reflexes for *\*-bódi* ‘goat’

<b>Jita</b>	imbusi / dʒimbusi	
<b>Kuria</b>	imburi / chimburi	
<b>Ngoreme</b>	emburi	emborete
<b>Ikoma</b>	amburi	amborete / chamborete
<b>Isenye</b>		emborete / chemborete
<b>Nata</b>		amborete / chamborete
<b>Ikizu</b>	imburi / zimburi	
<b>Zanaki</b>	emburi / ezimburi	

In this example we can distinguish Ngoreme, Ikoma, Isenye, and Nata as having slightly different wordforms with the *-ete* ending (although Ngoreme and Ikoma also have the additional form). Ikizu and Zanaki can be distinguished by having /z/ in the plural form, and Jita by the /s/ in the root. Of course, if we only look at one or two words, these differences don’t mean much, but if we look at hundreds of words and notice patterns that appear over and over again, then we can begin to make hypotheses about language groupings. Overall, however, Table 3 below shows us the lexical similarity percentages for the Mara Bantu languages. The highest similarities are between Ikoma and Nata (89%) followed by Isenye and Nata (88%), while the lowest are between Jita and Kuria (40%) and Jita and Ngoreme (42%).

Table 3. Lexical similarity percentages for Mara Bantu (Hill et al. 2007)

<b>Jita</b>							
40%	<b>Kuria</b>						
42%	73%	<b>Ngoreme</b>					
44%	62%	77%	<b>Ikoma</b>				
44%	61%	78%	85%	<b>Isenye</b>			
47%	63%	79%	89%	88%	<b>Nata</b>		
50%	62%	75%	77%	77%	84%	<b>Ikizu</b>	
50%	66%	76%	74%	74%	79%	87%	<b>Zanaki</b>

These lexicostatistics give us some justification for the groupings we saw in Figure 1 with Suguti languages (including Jita) separate from the rest of Mara proper. These percentages also lend some support to the split between *North Mara* (e.g. Kuria) and *South Mara* (Figure 2) groupings.

In terms of vowels, the vowel systems in Bantu are quite varied, but in Tanzania in particular the systems are usually restricted to 5- or 7-vowel systems. The typical 5-vowel system looks like the following:

i      u  
e      o  
a

Swahili is an example of this type of language, as is Jita, the sole example from the Shetler corpus. The 7-vowel systems can be of two different types:

i      u                      i      u  
**e      o                      ɪ      ʊ**  
ε      ɔ                      ε      ɔ  
a                                      a

The difference between these two systems is highlighted in bold and concerns the quality of the vowels at the second height level. (Note the lower-mid vowels are both /ε/ and /ɔ/ instead of /e/ and /o/. There are some 5-vowel systems which also have this type). The majority of Bantu languages in the Shetler corpus have the 7-vowel system on the left, including Ikoma, Ikizu, Isenye, Kuria, Nata, Ngoreme, and Temi. Zanaki is distinct in that it actually has the ten-vowel system represented below:

i      u  
**ɪ      ʊ**  
e      o  
ε      **ɜ**      ɔ  
a

The consonant systems in Mara Bantu resemble that of Swahili, as seen in Table 4.

*Table 4. Consonant phonemes, Swahili vs. Mara Bantu*

	<b>Swahili</b>	<b>Western Serengeti</b>	<b>Ikizu</b>	<b>Zanaki</b>
Stop	<i>p, t, k</i> <i>b, d, g</i> <i>tʃ, dʒ</i>	<i>t, k</i>  <i>tʃ</i>	<i>p, t, k</i> <i>b, d, g</i> <i>tʃ, dʒ</i>	<i>t, k</i> <i>g</i>
Fricative	<i>f, s, ʃ, h</i> <i>v, z</i>	<i>s, ʃ, h</i> <i>β, γ</i>	<i>s, ʃ, h</i> <i>β, z</i>	<i>s, h</i> <i>β, z</i>
Trill		<i>(r)</i>		
Flap	<i>ɾ</i>	<i>ɾ</i>	<i>ɾ</i>	<i>ɾ</i>
Glide	<i>l, w, j</i>	<i>w, j</i>	<i>w, j</i>	<i>w, j</i>
Nasal	<i>m, n, ɲ, ŋ</i>	<i>m, n, ɲ, ŋ</i>	<i>m, n, ɲ, ŋ</i>	<i>m, n, ɲ, ŋ</i>

The few major exceptions include the voiced fricatives /β/ and /γ/ which are often used instead of the voiced stops /b/ and /g/, especially in Western Serengeti, and the presence of /f/ and /v/ in Swahili (although Swahili loanwords into Mara Bantu can frequently carry [f] and [v] with them).

The Mara Bantu languages also make extensive use of lexical and grammatical tone, noun class prefixes, verbal morphology, focus, and other typical Bantu paradigms/structures. Please consult the Mara linguistic bibliography and especially Aunio et al. (2019) for more information.