Studies in African
Languages and Cultures, Vol. 54, 2020
ISSN 2545-2134; e-ISSN 2657-4187
DOI: 10.32690/SALC54.2

Rasmus Bernander
Antti O. Laine
Lotta Aunio
Department of Languages
University of Helsinki

The numeral system(s) in Western Serengeti: Formal, functional, and historical inferences

Abstract

In this study we offer a detailed synchronic and diachronic account of the hitherto un(der)-documented numeral systems of the four closely related (Eastern) Bantu language varieties: Ikoma, Nata, Ishenyi, and Ngoreme – together forming the Western Serengeti subgroup. We describe the essentially identical formation and organization of numerals in these language varieties while also noting the morphosyntactic behaviour of numeral expressions and their extended uses. Based on an extensive quantity of comparative data, we furthermore disentangle the historical background to the numerals and their systematization in Western Serengeti, connecting this specific linguistic domain with the wider genealogical profile of this subgroup.

Keywords: Bantu, Mara, numerals, numeral system, reconstruction

1. Introduction

In this article[1] we focus on four closely related and poorly described (Eastern) Bantu language varieties acknowledged as forming a genealogical subgroup, i.e. the Western Serengeti (WS) branch of the Mara subgroup of Great Lakes languages (Gibson & Roth 2019, Schoenbrun 1990). Three of the four language varieties constituting the WS group are Ikoma (ca. 15 000 speakers), Nata (ca. 11 500 speakers), and Ishenyi (ca. 9 500 speakers), jointly classified with the iso-code *ntk* and Guthrie code JE45[2]; the final language variety is Ngoreme (ca. 55 000 speakers), classified as *nqk* and JE401, respectively (see Aunio et al. 2019). They are all spoken in the Serengeti district of the Mara region, North-Western Tanzania, an area situated between Lake Victoria to the west and the Serengeti National Park to the east. The Mara region is dense and diversified from a linguistic perspective. Including the members of the WS group, it consists altogether of roughly 20 Bantu language varieties, along with other languages of Nilotic descent.

This study accounts for the numeral system(s) of the WS group languages. Numeral systems have clear semantic delimitations and morphosyntactic behaviour (Rischel 1997, Hammarström 2010), and even in more rudimentary descriptions of Bantu languages, numbers are typically still included. Numerals in Bantu are a formally distinctive class with a separate type of agreement marking (Stappers 1965). Consequently, there is plenty of early historical-comparative works in Bantuistics which pay attention to the origin of numerals and the developments within numeral systems in Bantu languages (see *inter alia* Werner 1919: 133-143, Schmidl 1915, Meinhof 1948: 117-124, Meeussen 1967: 96-98, 105, 117, Meeussen 1969, Hoffmann 1953, Polak-Bynon 1965, Stappers 1965). However, numerals are seldom a topic in modern comparative Bantu studies (an exception is Pozdniakov 2018, which, however, is a broader study encompassing the whole (putative) Niger-Congo macro family)[3]. This is surprising, given the fact that the "domain of numerals presents a prime case of using structured groups of lexemes for assessing historical-comparative questions" (Güldemann 2018: 74). Similarly, descriptive works on single Bantu languages or on small subsets of languages usually do not offer reconstructions of the origins of numerals and their further evolution within the numeral system.

---

[1] This research has been funded by the Kone Foundation. We wish to gratefully acknowledge their support. We also wish to thank our Western Serengeti language consultants, the Mara branch of SIL International and Tim Roth. We also thank Mary Chambers for polishing our English.

[2] See subsection 3.1 for more information about this Bantu specific referential system.

[3] See also Grimm (2019) for a critical review of this work.

Following the appeal by Rankin (2006), this study sets out to show how a historical-comparative approach is a particularly useful tool for the analysis and description of previously un(der)described varieties. Placing the WS numerals in a historical-comparative framework, this study brings further light to the diachronic forces behind the system, consequently offering a more robust description of it. At the same time, describing the numeral systems of WS also provides extra data for further comparative (and typological) work, facilitating the drawing of more fine-grained generalizations and conclusions about this linguistic notion. To this we may add that the documentation of numeral systems in these Tanzanian language varieties is a particularly pressing matter, insofar as they are increasingly being replaced by numerals borrowed from Swahili – that is, the prominent language which is both the national and co-official medium of communication (see e.g. Legère 2006) – and thus are at imminent risk of disappearing. This loss is a situation they share with many numeral systems of the world. Thus, Comrie (2005a) considers numeral systems of the world and their socio-cultural particularities a specifically endangered domain of languages.

In a fashion congenial to the bifocal descriptive-*cum*-comparative aim of this study – and in accordance with Blažek's (1999) three steps of numeral analysis – the article is organized in the following manner. In section 2 we present the overall numeral system with regard to the organization of ordinal and cardinal numerals, their agreement marking, and other formal and functional traits. In section 3 we focus on the historical-comparative background to this numeral system. In section 4 we offer a summary and some final conclusions.

2. Presenting the numeral system

In this study we define numerals, following Hammarström (2010: 11, see also Schapper & Klamer 2014), as "spoken normed expressions that are used to denote the exact number of objects for an open class of objects in an open class of social situations with the whole speech community in question"[4]. We furthermore treat these numeral expressions as being systematically arranged into a numeral system. As pointed out by Rischel (1997), numeral systems form a closed and relatively limited functional-semantic domain. That is to say, although an enormous number of different digits may indeed be formed in

[4] Notice that by following this definition we exclude "inexact" numerals for this study, like the reflexes of the common Bantu quantifiers (see Zerbian & Krifka 2008) or different fractions ('half', 'quarter' etc.).

a language, there is only a limited and closed set of primitive (mostly low-valued) numerals, of which all other digits are merely complex derivatives. The numeral system is typically further subdivided into cardinals and ordinals and we follow that convention here when describing the numeral systems of WS. However, as will be further evidenced in this section, there are also several other functional traits associated with numerals in these language varieties.

2.1. Cardinals

The simplex or basic cardinal numbers, as they occur in the WS language varieties, are presented in Table 1[5]. Note that there is no dedicated expression for 'zero', nor does any such expression occur in the formation of higher digits either. This is in line with the typical case in natural languages (see Greenberg 1978, 2000, Hurford 1987: 95).

TABLE 1. Cardinal numerals in WS

	Ikoma	Nata	Ishenyi	Ngoreme
1	-mu	-mwɛ	-mwe	-mwe
2	-βere	-βere	-βere	-βere
3	-tato	-tato	-tato	-tato
4	-ne	-nɛ	-ne	-ne
5	-taano	-taano	-taano	-taano
6	-saasaβe	-saasaβe	-saasaβe	-saansaβa
7	mohuŋgáte	mohuŋgáte	muhuŋgáte	mohúŋgate
8	-naane	-naanɛ	-naane	-naane

[5] The WS examples cited in this paper are from a corpus of transcribed and analyzed recordings made during extensive fieldwork in the Mara region from 2008 to 2019. Some of the data used has been collected by SIL members. Bible quotations referred to also come from the work by SIL. As Ikoma is the only WS language with an approved orthography, the writing system in the WS examples here is phonological (IPA), with the exception of contrastive long vowels, which are written with double consonants. The Ishenyi vowel system appears to be going through a loss of phonemic ATR contrasts, and there is a lot of inter- and intra-speaker variation. To keep the data comparable across the WS varieties, Ishenyi is also transcribed with 7 vowels despite occasional inconsistencies (see Laine 2016). Only surface tones are marked (with an accent), and numeral stems are not marked for tone as tones can also be realized on the numeral prefix and not the stem. It should be noted that tone analysis for these languages is work in progress.

	Ikoma	Nata	Ishenyi	Ngoreme
9	kénde	kénde	kénde	kénda
10	ikómi	ikómi	ikomi	ikómi
100	eriiɣána	rííɣana	riiɣana	riiɣána
1000	eɣekú	eɣekwé	eɣekwe	eɣekwé
100 000	ekirára	ekirára	ekiráre	ekirára

As seen in Table 1, the various numeral systems are more or less identical. The exceptions are, apart from some minor differences in vowel quality associated with more general differences in phonological structure, the extra word-medial nasal in '6' in Ngoreme. Additionally, word-final syllables in Ikoma may not be labialized, which induces the differing shapes of the numerals '1' and '1000' in this variety. The similarity between the varieties is not surprising, given the fact that they are closely related, with an estimated lexical overlap of up to 85% between Ikoma, Nata, and Ishenyi and 77% between Ngoreme and Ikoma (Roth 2018: 12).

As further indicated in this table, the cardinal numeral system consists of both variable and invariable numerals. Compare example (1), where the numeral '3' agrees with the head noun, with (2), where the numeral '9' remains unaffected.

(1) Ngoreme

ni-sɔ́m-ire	*e-B5i-táβo*	*βi-táto*
SP1SG-read-PFV	AUG-8-book	8-three

'I have read three books.'

(2) Ikoma

a-βá-áto	*kénde*	*m-ba-aré*	*βá-raru*	*a-mo-ɣóndo*
AUG-2-person	nine	FOC-SP2-COP.PST	SP2-leave	25-3-farm

'Nine people were leaving the farm.'

Bantu languages are characterized by having an extensive gender-like system, consisting of up to 20 noun classes, of which most are paired based on number (singular/plural). The WS language varieties form no exceptions to this characteristic trait. The variable numerals may take agreement with most of these different noun classes as they occur among the WS members. We use data from Nata to illustrate this fact in Table 2. (See also Table 3 in subsection 3.2).

TABLE 2. The enumerative agreement prefixes in Nata

Num. / NC pair	1/2 u-/βa-	3/4 u-/e-	5/6 ri-/a-	7/8 ki-/βi-	9/10 i-/i-	12/13 ka-/tu-	14 βu-
'1' **-mwɛ**	ú-mwɛ	ú-mwɛ	rí-mwɛ	kí-mwɛ	í-mwɛ	ká-mwe	βú-mwɛ
'2' **-βere**	βá-βere	é-βere	á-βere	βí-βere	í-βere	tú-βere	βó-βere
'3' **-tato**	βá-tato	é-tatɔ	á-tato	βí-tato	í-sato	tú-tato	βó-tato
'4' **-nɛ**	βá-nɛ	é-nɛ	á-nɛ	βí-nɛ	í-ɲɛ	tú-nɛ	βú-nɛ
'5' **-taano**	βá-taano	é-taanɔ	á-taano	βí-taano	í-saano	tú-taano	βú-taano
'6' **-saasaβe**	βa-saasáβe	e-saasáβe	a-saasáβe	βi-saasáβe	i-saasáβe	tu-saasáβe	βu-saasáβe
'8' **-naanɛ**	βa-naanɛ́	e-naanɛ́	a-naanɛ́	βi-naanɛ́	i-ɲaaɲɛ́	tu-naanɛ́	βu-naanɛ́

All the other WS members behave in essentially the same way as Nata when it comes to numeral agreement. The Ikoma and Nata enumerative prefixes have the close back vowel /u/ for the numeral '1', and Ishenyi and Ngoreme the close mid back vowel /o/. Underlyingly, the vowel in Ikoma and Nata may also be the close mid vowel which is dissimilated from the −ATR stem vowel /ɛ/. In Nata, the −ATR vowel is still present in the stem *-mwɛ*, whereas Ikoma (as mentioned above) has lost the final vowel due to a later rule that prohibits labialization from occurring word-finally.

One exceptional feature of Nata is the regular singular-plural shift of the diminutive noun class 12 *ka-* to class 13 *tu-*, as in (3). The other WS members derive plurals of class 12 *ka-* with the prefix of class 19 *hi-* (another noun class dedicated to diminutives), as in (4).

(3) Nata

a-*ka-té*	*ké-mwe,*	*o-tu-té*	*tú-taano*
AUG-12-tree	12-one	AUG-13-tree	13-five

'one small tree, five small trees'

(4) Ikoma

a-ka-té	*ká-mu,*	*e-hi-té*	*hí-taano*
AUG-12-tree	12-one	AUG-19-tree	19-five

'one small tree, five small trees'

The enumerative class prefixes for classes 11 *ru-*, 15 *ku-*, 16 *ha-*, and 20 *ɣu-* are not included in this table as they only agree with the numeral '1'. Curiously, class 14 is also a singular class. However, head nouns in this class can still be modified with plural numerals, as evident in (5).

(5) Ishenyi

oβó-óŋgo	*m-bu-tato*	*áá-na*
14-brain	FOC-14-three	SP1-com

'S/he is not very intelligent.' (Lit: 'S/he has three brains.')

Similarly, all WS members lack enumerative prefixes for the locative noun classes 17 and 18 (as well as for the almost completely obsolete locative noun class 25, cf. Grégoire 1975: 170-175). Consequently, they are not included in the table either. In fact, locative class agreement on numerals is generally rare given two conspiring facts. Firstly, the locative noun classes are overwhelmingly devoid of any inherent nouns. Instead, they are applied onto nouns of other noun classes to mark notions of location. Secondly, there is a general restriction in the WS which stipulates that any modifiers of such a noun derived with an additive locative class prefix do not agree with the locative but with the lexical noun class (Aunio et al. 2019). The only exception to this pattern, with locative enumerative agreement marking, occurs when the class 16 enumerative prefix is employed for marking agreement with either of the two nouns inherently belonging to class 16, that is, *ahasé* and *ahaɣíro*, both meaning 'place'. This is illustrated in (6).

(6) Ishenyi

a-ha-sé	*ha-mwe*	*há-no*	*ha=há-no*	*ha=há-áne*
AUG-16-place	16-one	16-DEM.DIST	SP16=16-DEM.PROX	SP16=16-POSS.1SG

'The one place which is there is mine.'

As is typical for Bantu languages (see e.g. Schadeberg 2003: 150), the enumerative agreement prefixes form a distinct paradigm of agreement markers. These prefixes, as represented in Table 2, are largely identical to the set of pronominal prefixes used for other nominal modifiers. The main differences are found in classes 4 (*e-*), 6 (*a-*), and 10 (*i-*), which lack an initial consonant in the enumerative form, unlike in the pronominal forms (which are *ɣe-*, *ɣa-*, and *t͡ʃe-*, respectively).

We can illustrate this fact with example (7) from Ishenyi, where the same head noun, derived in noun class 10, triggers agreement in the noun phrase which is realized differently on the possessive pronoun than on the numeral. In (8) from Ikoma, it is the demonstrative and connective which agree differently from the numeral, all agreeing with the noun *amaɲémbe* 'mangoes' of noun class 6.

(7) Ishenyi

βo-ɣ-íkar-a	*t͡ʃe-síku*	*t͡ʃ-et͡ʃe*	*i-ʃáto*
14-IPFV-stay-FV	10-day	10-POSS.3SG	10-three

'It will stay for three days' (lit. It will stay its three days).'

(8) Ikoma

a-ra-mú-saβ-a	*a-mu-h-ɛ́*	*a-ma-ɲémbe*	*ɣá-jo*	*ɣa*	*ma-ɣúta,*
1-PROG-OP1-pray-FV	SP1-OP1-give-SBJV	AUG-6-mango	6-DEM.REF	CONN6	6-oil
o-mo-βúréní	*a-ra-mú-h-a*	*a-ma-ɲémbe*	*a-tato*		
AUG-1-young_man	SP1-PROG-OP1-give-FV	AUG-6-mango	6-three		

'He asked him to give him those oily mangoes, the young man gave him three mangoes.'

As also illustrated in this example and as further seen in Table 2, the enumerative prefix of class 10, constituted by a single high /i/, affects the phonological make-up of some of the numerals. The initial /t/ in '3' and '5' is fricativized to either /s/ (Ngoreme and Nata) or /ʃ/ (Ikoma and Ishenyi). The initial /s/ of '6' also shifts to /ʃ/ in Ikoma and Ishenyi. The initial nasal of '4' shifts to a palatalized /ɲ/. Interestingly, both nasals occurring in '8' are palatalized in this manner. As further discussed in section 4, these morphophonological alternations come with historical implications.

Regarding the invariable numerals, they are formed as nouns and are assigned to different noun classes. Only '7' belongs to noun class 3; '9', '10', and '100' instead belong to class 5/6. The highest basic digits of the system, '1000' and '100 000', belong to class 7/8. The noun class prefix of class 5 behaves irregularly with numerals associated with this class, unlike other nominal stems which regularly take a full CV shaped prefix *ri(i)-*. The numeral *rii-ɣána* '100' keeps this form, but for *i-kómi* '10' this prefix is reduced to *i-* and for *kénde* '9' the prefix is further reduced to zero. (But see subsection 2.2 where it is shown that the full prefix re-appears in ordinal constructions in Ishenyi.)

Bantu languages are generally considered as having decimal-based numeral systems, given the fact that the base – that is, the numerical value used recursively to form other numerals – is a stem meaning 'ten'. Hence, the numerals 11-19, as well as the decades, are typically synchronically transparent complex con-

structions derived with 'ten'[6]. This system adheres to the pattern which is also by far the most common from a cross-linguistic perspective. As Comrie (2005b) notes: "We live in a decimal world". Thus, not surprisingly, the members of the WS group also have decimal-based numeral systems. Notice that, with the exception of the decimal-based system, there is also a recurrent pattern across the Bantu speaking area indicative of a quinary or base-five system (for digits below 10). However, although some close relatives/neighbours have such a quinary system (mixed with a decimal system), this is not the case in the WS language varieties (see subsection 3.4 for a more elaborate account).

Greenberg (1978, 2000, see also Comrie 2005a, 2005b, Schapper & Klamer 2014) distinguishes between "additive" and "multiplicature" complex numeral constructions, describing the most common arithmetic operations applied to the base and other numeral components in the formation of numerals. The WS language varieties form their numerals 11-19 through an additive strategy. As apparent in Table 1, the word for the (cardinal) numeral '10' in all the WS languages is *ikómi*, which also forms the augend (i.e. the count base) in these case(s), the addend being any number in the series from 1-9. The comitative preposition/conjunction na 'and/with' functions as the "link", that is, the additive operator between the augend *ikómi* and the serialized addend. The full set of these numerals are illustrated in (9) from Ikoma, with class 1 and 2 agreement.

(9) Ikoma

ikómi na ú-mu '11'
ikómi na βá-βere '12'
ikómi na βá-tato '13'
ikómi na βá-ne '14'
ikómi na βá-taano '15'
ikómi na βa-saasáβe '16'
ikómi na mohuŋgáte '17'
ikómi na βa-naané '18'
ikómi na kénde '19'

Notice that the final numeral in these examples agrees with the head noun and not with the base, as evidenced in (10). Example (11) from Ngoreme furthermore illustrates the analysability of the building blocks of these numerals. As seen at the end of this sentence, the augend base is used only once to cover a series of

[6] However, Hammarström (2010) mentions some interesting deviations from this pattern, particularly in North-Western Bantu.

numerals and the comitative has been changed to a disjunctive coordinator 'or' (*au*, borrowed from Swahili) to signal the possible range of variation.

(10) Ishenyi

e-βi-ɣésɔ	*i-kómi na*	*βi-βere*
AUG-8-knife	5-ten COM	8-two

'12 knives'

(11) Ngoreme

o-mó-óna	*we*	*ɣe-súβe*	*a-hík-i*	*i-kómi*	*na*	*e-nááne*
AUG-1-child	CONN1	7-male	SP1-arrive-CAUS	5-ten	COM	7-eight

we	*ki-ɣáíkoro*	*a-híki*	*e-mé-ka*	*e-kómi na*	*e-táto*	*au e-né*	*au e-tááno*
CONN1	7-female	SP1-arrive-CAUS	AUG-4-year	4-ten COM	4-three	or 4-four	or 4-five

'A male child reaches 18 [years] and a female [child] reaches 13, 14, or 15 years.'

Decades, that is, the numerals '20'-'90', are also decimal-based. However, they are formed differently in two aspects. Firstly, and crucially, they are not formed with *ikómi* but with the alternative word *mirɔ́ŋgɔ* (*mu-rɔ́ŋgɔ* in the singular, hence belonging to noun classes 3/4). Furthermore, they are formed through multiplication, that is, constructions where the decimal base (in this case *mirɔ́ŋgɔ*) serves as a multiplicand and any number from 2-9 may serve as a multiplier. The relationship between the two components is marked via noun class agreement governed by the multiplier. Example (12) illustrates this with data from Ishenyi.

(12) Ishenyi

mirɔ́ŋgɔ eβere '20'
mirɔ́ŋgɔ etáto '30'
mirɔ́ŋgɔ eene '40'
mirɔ́ŋgɔ etaanó '50'
mirɔ́ŋgɔ esaasáβe '60'
mirɔ́ŋgɔ muhuŋgáte '70'
mirɔ́ŋgɔ enaané '80'
mirɔ́ŋgɔ kénde '90'

Numerals within these decades are formed with the combination of the multiplicative strategy and the additive strategy described for numerals 11-19. This is illustrated in (13) and (14) below. Notice again that the final numeral in these examples agrees with the head noun and not with the base.

(13) Ikoma

a-βá-áto	*mi-rɔ́ŋgɔ*	*e-βere na*	*βá-taano*
AUG-2-person	4-ten	4-two COM	2-five

'25 people'

(14) Ikoma

n-t͡ʃe-eɲí-ho	*t͡ʃa-ŋɔ́mbɛ*	*i-reŋgé?*	*t͡ʃe-eɲi-hó*	*mi-rɔ́ŋgɔ*	*e-βere n-í-ʃaasáβe.*
FOC-SP10-COP-LOC	10-cow	5-how_many	SP10-COP-LOC	4-ten	4-two COM-10-six

'How many cows are there? There are 26.'

Schadeberg (2003: 150) subdivides the Bantu cardinals into two further subtypes, that is, "referential" and "absolute" numerals, respectively. Whereas referential numerals – constituted by all examples provided up to now – are used for counting individuals or entities ('one/two/three X'), absolute numerals are dedicated to calculations ('one, two, three…'). Absolute numerals are always inflected in class 9/10 in all four varieties, for example, Ngoreme *emwé* '1', *iβére* '2', *isáto* '3'. In contrast with a common trait in other Bantu languages (Vanhoudt 1994), there is no difference in the formal realization of the word stem '1' when used as an absolutive relative to its use as a referential. One potential exception is found in Ikoma, where the final vowel in the stem for '1' fluctuates between /u/ and /a/ in the formation of '11' and other additive numeral constructions with '1' as the addend. Compare (9) above with *aβáána βaat͡ʃe ikómi na úmwa* 'his eleven sons' (Genesis 32:22).

The forms used for expressing '100', '1000', and '100 000' are also included in Table 1 as they are simplex numerals which also serve as bases for other complex numeral derivatives[7]. The use of both '100' and '1000' is illustrated in (15) below.

(15) Ikoma

a-β-ikwaβé	*βa-ra-réu*	*βa-ra-mísi,*	*βa-ra-β-íít-a,*
AUG-2-MAASAI	SP2-PROG-get_drunk	SP2-PROG-sleep	SP2-PROG-OP2-kill-FV

βa-y-iit-a	*a-ma-yána,*	*a-ma-yána*	*mpaká*	*akuβá*	*e-ye-kú*	*ki-mu*
SP2-NAR-kill-FV	AUG-6-hundred	AUG-6-hundred	until	maybe	AUG-7-thousand	7-one

'The Maasai got drunk and slept, then they killed them, they killed hundreds, hundreds up to maybe a thousand.'

---

[7] We follow the criteria by Schapper & Klamer (2014) that the lowest recursively occurring base designates the system. Hence, as this is '10' in the WS group, the numeral systems of these languages are to be treated as decimal-based.

Notice that the disparate formal realization of '1000' in Ikoma, as seen in this example, is due to the same phonological phenomenon as discussed for '1' above (intrinsically illustrated in this example as well). Thus, what in the other varieties is pronounced with a glide, *ɣe-kwé*, is pronounced without word-final labialization in Ikoma.

These higher digits may also be multiplied and serialized as bases using a strategy parallel to that described for forming decades.

(16) Ishenyi

ama-ɣána a-táto '300'
e-βe-kwé βii-ne '4000'
e-βi-ráre βi-taanó e-βe-kwé miróŋgɔ e-saasáβe na muhuŋgáte na ama-ɣána a-naané na miróŋgɔ kénde na i-βere '567 892'

2.2. Ordinals

As is common across the Bantu speaking area (see Van de Velde 2013, 2019), ordinals are formed as "numeral possessives" (Schadeberg 2003: 150), that is, in a complex construction consisting of a head noun and the connective followed by the numeral[8]. This construction is illustrated in (17) to (19) below.

(17) Ishenyi

o-mó-ónto	*o*	*rii-kómi*
AUG-1-person	CONN1	5-ten

'the tenth person'

(18) Ngoreme

neʧu	*a-kw-ínɔk-a*	*o-ro-síko*	*ro*	*ɣa-táto,*	*ro*	*ká-ne,*	*ro*	*ɣa-táno*
thus	SP1-SIT-depart-FV	AUG-11-day	CONN11	12-three	CONN11	12-four	CONN11	12-five

'Thus, while s/he left [to go to the farm], on the third day, the fourth [day], the fifth [day]…'

(19) Ikoma

a-rá-áʧ-a	*o-mo-βúréní*	*o-wó-nde*	*ó*	*ka-βere […]*
SP1-PROG-come-FV	AUG-1-young_man	AUG-1-other	CONN1	12-two

[8] The connective marker is commonly used in Bantu languages to connect a head noun with another modifying nominal constituent. The connective is formally different in the WS group from the canonical Bantu reflex *-a* (see Aunio et al. 2019: 517-518 for further details).

a-rá-átʃ-a	*o-mo-βúréní*	*o-wó-nde*	*ó*	*ɣa-tato […]*
SP1-PROG-come-FV	AUG-1-young_man	AUG-1-other	CONN1	12-three
a-rá-átʃ-a	*o-mo-βúréní*	*o-wó-nde*	*ó*	*ka-ne*
SP1-PROG-come-FV	AUG-1-young_man	AUG-1-other	CONN1	12-four

'Then came another (second) young man […] Then came another (third) young man […] Then came another (fourth) young man.'

Several facts regarding the ordinal construction may be deduced from these examples. Firstly, we may note from example (19) that other modifiers may interfere between the head noun and the connective construction containing the ordinal within the noun phrase. Secondly, in (17), we see that the full CV-shaped noun prefix of class 5, which is otherwise reduced to /i-/ in *ikómi* '10', re-occurs within ordinal connective constructions in Ishenyi. Thirdly, in examples (18) and (19) we see that, in addition to the connective, the numeral is inflected with a prefix *ka-/ɣa-*, the variation in consonant realization being conditioned by Dahl's Law[9]. However, if the numeral is the invariable '7','9', or '10', this prefix may not surface. Thus, compare the different realizations of '5', '6', and '7' in (20). The omission of *ka-/ɣa-* with invariable numerals is further illustrated in (20) and (21).

(20) Ikoma

o-ra-már-a	*ɣo-ɣí-karaŋg-a, ró*	*ɣa-taano, ró*	*ɣa-sasááβe, ro*	*mohuŋgáte*
SP2SG-SIT-finish-FV	INF-OP7-fry-FV	CONN11 12-five	CONN11 12-six	CONN11 seven

'When you finish frying it the fifth [day], the sixth or the seventh'

(21) Ishenyi

e-híti	*e*	*muhuŋgáte*
9-hyena	CONN9	seven

'the seventh hyena'

Similar to what has been pointed out by, for example, Stappers (1965) for several other (Eastern) Bantu languages, the prefix *ka-/ɣa-* is also used for deriving multiplicatives of numerals in the WS language varieties. This is illustrated in (22) and (23).

---

[9] Dahl's law refers to a type of dissimilation process common in North-Eastern Bantu, where voiceless stops are voiced when the succeeding syllable also consists of a voiceless consonant. See Davy & Nurse (1982) about this feature in general, and Aunio et al. (2019) for the Mara languages in particular.

(22) Ikoma

ni-sɔm-ire	*e-ɣi-táβu*	*kí-no*	*ká-βere*
SP1SG-read-PFV	AUG-7-book	7-DEM.PROX	12-two

'I have read this book twice.'

(23) Ishenyi

ne-e-mu-kúmbat-irɛ	*ɣa-táto*
SP1SG-PST-OP1-embrace-PFV	12-three

'I embraced him three times.'

The ordinal 'first' constitutes the only exception to the pattern in the WS group where ordinals are directly derived from cardinals. As is common cross-linguistically and in Africa in general (Stolz & Veselinova 2013), as well as in Bantu in particular (Polak-Bynon 1965), 'first' is instead formed through suppletives, that is, derivationally independent forms. Whereas Ikoma, Ishenyi, and Nata form the ordinal 'first' exclusively with the adjective *-mbɛrɛ*, as evidenced in (24) to (26), Ngoreme forms it with either *-mbɛrɛ* (27) or *kwánsa* (28). Whereas *-mbɛrɛ* acts as an adjective and takes regular nominal agreement, *kwánsa* is formed with the use of the connective marker.

(24) Ikoma

a-βá-á͡ʧokoro	*βa-ane*	*a-βá-mbɛrɛ*
AUG-2-grandchild	2-POSS1SG	AUG-2-first

'my first grandchildren'

(25) Ishenyi

rii-βáɣa	*e-rí-mbɛrɛ*
5-time	AUG-5-first

'the first time'

(26) Nata

rii-βurúŋga	*e-rí-mbɛrɛ*
5-egg	AUG-5-first

'the first egg'

(27) Ngoreme

a-βá-nto	*βa-no*	*βa-ɲúma*	*βóóno,*	*βa-ɾá-β-a*	*βá-mbɛrɛ,*
AUG-2-person	2-DEM.PROX	2-behind	now	SP2-IPFV-be-FV	2-first
na	*βá-mbɛrɛ*	*βóóno,*	*βa-ɾá-β-a*	*βa-ɲúma*	
COM	2-first	now	SP2-IPFV-be-FV	2-behind	

'There are those who are last who will be first, and first who will be last.' [Luke 13:30]

(28) Ngoreme

e-tjáɲi	*ja*	*kwánsa*	*e-ka-hét-a*
9-animal	CONN9	first	SP9-NAR-pass-FV

'The first animal passed.'

2.3. Functional traits: Morphosyntactic behaviour and extended uses

Insofar as the numerals have a particular set of agreement prefixes, they may be morphologically defined as constituting a word category of their own (cf. Schadeberg 2003, Stappers 1965, see also Greenberg 2000). However, as pointed out in subsection 2.1 above, some numerals do not inflect for agreement. Moreover, their morphosyntactic behaviour is in many ways identical to that of other adnominal modifiers. Cardinals behave identically to adjectives (see Van de Velde 2019) and ordinals may be subsumed within a more general framework of connective constructions (as touched upon already in subsection 2.2). Thus, in accordance with the general head-driven typological structure characterizing Bantu languages, where modifiers tend to follow the head they modify (Van de Velde 2019), the numeral typically follows the noun it modifies in the WS language varieties (as may be deduced from all the previous examples in this article). With that said, however, a numeral seems to be allowed to occur relatively freely in a clause. As evident in the various realizations of the same proposition in (29), a numeral can even precede the head noun.

(29) Ikoma

t͡ʃa-séése t͡ʃé-éne in-kóro í-ʃato
10-dog 10-POSS1SG 10-big 10-three
or *t͡ʃa-séése t͡ʃé-éne í-ʃato in-kóro*
10-dog 10-POSS1SG 10-three 10-big
or *i-ʃato t͡ʃa-séése t͡ʃé-éne in-kóro*
10-three 10-dog 10-POSS1SG 10-big
'three big dogs of mine'

The pragmatic-semantic explanations that might underlie such alternations require further research.

Another feature in need of further investigation is the fact that variable numerals generally do not take the augment in WS. The augment is a functionally elusive nominal pre-prefix whose presence is dependent on a number of factors often connected to notions such as specificity, topicality, and definiteness (at least diachronically, see de Blois 1970, Van de Velde 2019: 249-254). In other Bantu

languages, a numeral automatically carries an augment if the governing noun has one (cf. de Blois 1970). Furthermore, just as with adjectives, it is common to add the augment on numerals to make them nominalized and non-restrictive (Van de Velde 2019: 262-263). However, in neither of these two contexts does an augment occur in the WS language varieties: the former is evident in (1) above, just to mention one of several examples in the paper illustrating this fact; and the latter is illustrated in (30) below. Note that the invariable nominalized numerals do not adhere to these restrictions but take the augment like any other noun, as evident for example with *a-ma-ɣána* 'hundreds' and *e-ɣe-kú* 'thousand' in (15) above (but where *ki-mu* '1' modifying 'thousand' occurs without an augment).

(30) Ikoma

βa-taano m-ba-mar-iri ɣo-kú, m-ba-saaɣ-iri βa-βere, u-mu n-a-aɣu-káŋgat-a
2-five FOC-2-finish-PFV INF-die FOC-2-remain-PFV 2-two 1-one FOC-3SG-IPFV-lead-FV
'(among these kings) five have fallen, two remain, one is (=reigns)…' [Revelation 17:10]

Apart from such morphosyntactic processes, numerals may be used in extended ways to express specific functional traits. Distributive numerals, namely numerals marking a multiple set of a specified number (see e.g. Gil 2013, Greenberg 2000) are derived from cardinals through reduplication. This is illustrated in (31), where (31a) indicates that the participants carried two cups in total, whereas (31b), that is, the distributive numeral construction, indicates that the participants brought with them two cups each; hence, in total four cups.

(31) Ishenyi

a) *m-ba-ɣɛɣ-irɛ* *e-βi-kɔ́mbɛ* *βi-βere*
FOC-SP2-carry-PFV AUG-8-cup 8-two
'They carry 2 cups (in total).'

b) *m-ba-ɣɛɣ-irɛ* *e-βi-kɔ́mbɛ* *βi-βere* *βi-βere*
FOC-SP2-carry-PFV AUG-8-cup 8-two 8-two
'They carry 2 cups (each).'

Forming distributive numerals through reduplication is a typologically common strategy, including in the Bantu family, as shown by Gil (2013). According to this author, the motivation behind its ubiquity is iconicity, as the reduplication directly corresponds to the conceptualization of a multiple set of entities.

Another type of numeral derivative, also described for the related language Gusii (Cammenga 2002: 349-351), comes from the use of plural referent agreement with the ordinal '1'. As illustrated in (32) with examples from Ngoreme, such constructions have a partitive ("unit of") reading.

(32) Ngoreme

a) *t͡ʃi-ŋɔ́mbɛ* *t͡ʃi-no* *t͡ʃe-mwé*
10-cow 10-DEM.PROX 10-one
'this group of cows'

b) *a-βá-nto* *βa-mwé*
AUG-2-person 2-one
'a group of people'

c) *e-βi-táβo* *βi-no* *m=be-mwé*
AUG-8-book 8-DEM.PROX COP=8-one
'These books are of a particular type.'

When the ordinal '1' takes agreement with the locative noun class prefix 16 it functions as a reciprocal adverbial forming the meaning 'together'. This is illustrated with example (33) from Ishenyi.

(33) Ishenyi

e-ɣi-súsu *m-be-réŋge* *βe-ɣw-íkar-a* *ha-mwé* *n=in-ká*
AUG-7-rabbit FOC-SP8-COP.PST SP8-IPFV-live-FV 16-one COM=9-lion
'The rabbit lived together with the lion.'

Besides this meaning, however, this form has additional distinctive functions, namely as a modal adverbial expressing epistemic possibility (34) and as a disjunctive coordinator 'or' (35).

(34) Ishenyi

hamwe *ha-ká* *a-are,* *ne-ku-rór-a* *i-βásikeli* *e-e͡t͡ʃe*
perhaps 16-home SP1-COP SP1SG-IPFV-see-FV 9-bicycle 9-POSS3SG
'Perhaps s/he is at home, I see her/his bicycle.'

(35) Ikoma

n-o-oɣo-tór-a *ku-mísi* *há-no* *hamu* *u-ɣarúk-e* *a-ká*
FOC-SP2SG-IPFV-can-FV INF-sleep 16-DEM.PROX or SP2SG-return-SBJV 23-home
'You can sleep here, or you can return home.'

Finally, we may note that numerals are found in lexicalized constructions such as in compounds referring to the names of the day of the week, as illustrated in (36). Notice that the counting of days starts from Sunday (unlike in the co-official and national language Swahili, where the first day of the week is Saturday).

(36) Ishenyi

orusíku orumbɛrɛ 'Sunday'
orusíku roo kaβere 'Monday'
orusíku roo ɣatáto 'Tuesday'
orusíku roo kane 'Wednesday'
orusíku roo ɣataanó 'Thursday'
orusíku roo ɣasaasáβe 'Friday'
orusíku roo muhuŋgáte 'Saturday'

3. Historical-comparative implications

After having described the formal and distributional characteristics of the numerals in the WS language varieties, in this section we attempt to unravel the semasiological background of these forms and to account for the further historical implications that this gives rise to[10]. To facilitate this task, we first have to explain the classificatory and genealogical particularities of the WS group.

3.1. Classificatory profile and genealogical background of the WS group

The Bantu languages are most commonly classified through an alpha-numeric referential system first developed by Guthrie (1948, 1967-1971) and later updated by Maho (2003, 2009). In this system, consisting of zones (letters), further divided into groups (decimals) and individual languages (numbers), the Mara languages, to which the WS language varieties belong, were initially classified within zone E40, but were later reclassified together with several other languages spoken around Lake Victoria, previously belonging to zone D and E, into a zone J. *Pace* Maho (2003, cf. Philippson & Grollemund 2019), these languages are typically referred to with J as their first letter followed by their original classification, hence Mara = JE40.

The Guthrie classification system is primarily a geographical and not a genealogically based system. However, with the exception of JE41 Rogoori (which was re-assigned to the Luhya cluster), the Mara group or JE40 is often treated as a valid subgroup on genealogical grounds as well (see e.g. Nurse 1999). From a gene-

[10] Readers are referred to a table in the Appendix for references to sources on specific languages discussed in this part. The formal representation of the reconstructed forms presented in this section has been adapted in accordance with the orthographic conventions developed by the Tervuren group (see Schadeberg 2003: 147).

alogical standpoint, primarily based on lexico-statistics (see Nurse & Philippson 1980, Schoenbrun 1990, 1994, see also Hill et al. 2007)[11], the Mara branch has been categorized as belonging to the East Nyanza group, which is in turn a subgroup of the Great Lakes (GL) languages. For a visual representation of the genealogical relationship see Roth (2018: 111) (which is a slightly adapted version of that of Schoenbrun 1997: 12-13). The GL (also known as the (Inter)Lacustrine) group corresponds with zone J.

Two facts need to be highlighted that are of importance for the following discussion. Firstly, Nurse (1999: 27-28) points out that the GL languages on the eastern side of Lake Victoria, that is, the East Nyanza and (Greater) Luhya groups, are linguistically similar in a manner which is at the same time different from the other GL languages. This is surprising, as these groups are typically treated as being relatively distantly related. On the other hand, geographically they occupy contiguous areas, renowned for extended contacts. Secondly, as touched upon already in section 1, the area where the Mara languages, including the WS language varieties, are spoken is also characterized by contact with other non-Bantu linguistic communities (Ehret 1971, Shetler 2003: 11-14, 288, Dimmendaal 1995, Nurse 1999). In the present day this contact situation is primarily with South and West Nilotic languages (Datooga and Luo, respectively), but historically there has also been contact between ancestors of the WS group and Cushitic and earlier Nilotic linguistic communities (e.g. East Nilotic Maa).

The GL languages are, in turn, part of an Eastern Bantu group whose ancestors separated from their Western counterpart(s) somewhere in the Congo region roughly 2000 years ago (Grollemund et al. 2015). When referring to Eastern Bantu, we follow the latest phylogenetic classification of such a group, as provided by Grollemund et al. (2015), which, nonetheless, corresponds "fairly well" (Phillipson & Grollemund 2019: 346) with previous attempts at such a classification. We will differentiate between reconstructible shared material and patterns within Eastern Bantu and what is reconstructible, or has indeed been reconstructed, for Proto-Bantu, that is, the earliest ancestor of all Bantu languages.

Zooming in from the macro- to the micro-level, the Mara branch itself is divided into a South and North Mara (see Schoenbrun 1990). Gibson & Roth (2019) and Roth (2018: 110-111) argue for a further split of South Mara into a SW Mara subgroup – containing Ikizu and Zanaki – and the Western Serengeti, that is, the group of language varieties under consideration in this study.

---

[11] See also Nurse (1999) for a critique of Schoenbrun's selective methodology.

3.2. On the enumerative prefixes

Before any discussion of the actual numeral forms, a brief comment on the history of the enumerative agreement prefixes should be given. Schadeberg (2003: 150) notes that the reconstructions for these prefixes to Proto-Bantu are "somewhat shaky" and that there is typically interference with the set of pronominal and nominal prefixes. With that said, the WS members chiefly adhere to the reconstructed set of enumerative prefixes (Meeussen 1967: 97, see also Stappers 1965), as is made further evident in Table 3. The shakier reconstructions are indicated by question marks, in analogy with how they are represented in Meeussen's (1967: 97) reconstructions. Classes 17 and 18 are not included in this table as the WS language varieties have lost the reflexes of these enumerative prefixes (see also subsection 2.1). Conversely, the enumerative prefix of class 20, which exists throughout the Mara group – and is thus doubtlessly a shared retention in the WS subgroup – is not included in this table as it is not reconstructible for Proto-Bantu (see Maho 1999: 253).

An interesting exception, however, is the class 3 marker in Ngoreme, which has the velar onset (characteristic of the pronominal prefix) for numerals as well. As Schadeberg (2003: 150) considers the reconstruction of the class 3 enumerative prefix **ú-* as "less certain" (compared to that of class 10 for example), the question should be raised whether this is due to "interference" or whether it should be taken as counterevidence against the suggested reconstruction.

3.3. Numerals 1-5

For the reconstruction of the numerals, we start with the numerals 1-5, which are all simplex (monomorphemic) and variable stems in the WS language varieties. These numerals can all be straightforwardly connected with the forms reconstructed for Proto-Bantu (see Meeussen 1967: 105, Schadeberg 2003: 150), that is, **moi* '1', **bɪli* '2', **-tátʊ* '3', *-nai* '4', and *-táano* '5'.

In fact, all forms for the numerals 1-5, with the exception of 2, are further reconstructible for an assumed Proto-Niger Congo (see Pozdniakov 2018: 293, 313). Regarding the expression of the numeral '2', which is split between a western and an eastern form in Bantu, the WS languages pattern with the Eastern Bantu languages (see for example Guthrie 1961-1971, vol. III: 23) in having a reflex of **bɪli* and not **bali*. Furthermore, we may note that the WS varieties seem to have levelled out the formally distinctive absolute number **mʊ-oti* reconstructed for Proto-Bantu by Vanhoudt (1994), using the reflex of the referential numeral also for calculations.

TABLE 3. Enumerative prefixes in Proto-Bantu compared to the WS language varieties

Variety / 1 Noun class pair	1/2	3/4	5/6	7/8	9/10	11	12/13	14	15	16	19
Proto-Bantu	*ʊ? /*ba	*ʊ-? / *ɪ-?	*di- / *a-?	*kɪ- / *bi-	*ɪ-? / *i-	*dʊ-	*ka- / *tʊ-	*bʊ-	*kʊ-	*pa-	*pi-
Ikoma	u- / βa-	u- / e-	ɾi- / a-	ki- / βi-	i- / i-	ɾu-	ka-	βu-	ku-	ha-	hi-
Nata	u- / βa-	u- / e-	ɾi- / a-	ki- / βi-	i- / i-	ɾu-	ka- / tu-	βu-	ku-	ha-	-
Ishenyi	o- / βa-	o- / e-	ɾe- / a-	ke- / βi-	e- / i-	ɾo-	ka-	βo-	ko-	ha-	hi-
Ngoreme	o- / βa-	ɣo- / e-	ɾe / a-	ke- / βi-	e- / i-	ɾo-	ka-	βo-	ko-	ha-	hi-

Interestingly, it would seem that the conditioned lenition of the initial consonant of the numerals '3', '4', and '5', when occurring with the high fronted class 10 enumerative prefix in the WS varieties, is also a retention from Proto-Bantu. Thus, Meeussen (1967: 105) notes for Proto-Bantu (orthography slightly altered, see f.n. 9): "In class 10 the prefix has to be set up as *i-* […] with a peculiar representation in at least two stems: *icátu* 'three', *icáano* 'five' (and *inyai* 'four'?)". From a micro-comparative perspective, it is interesting to note the formal variation of the morphophonological realization of this specific feature as found across the East Nyanza languages, namely Mara and Suguti. In the Suguti languages (including Kwaya, which otherwise tends to pattern with the WS group, see e.g subsection 3.4), the reflexes of '4' and '5' do not alter their basic form when inflected with a class 10 prefix. On the other hand, the form for '3' has been reanalyzed as *satu*, that is, with a fricativized stem-initial consonant regardless of which agreement class prefix is in use (cf. Stappers 1965, Dimmendaal 2011: 57-58). Interestingly, there are two North Mara varieties that also pattern differently from the other Mara language varieties (see Aunio et al. 2019). Kabwa behaves similarly to Suguti, whereas in Simbiti, the class 10 fricativization is only optional with '3', i.e. *i-tatɔ* ~ *i-satɔ*.

3.4. Numerals 6-9

Compared with the numerals 1-5, which can be linked to Proto-Bantu reconstructions, the semasiological background of the numerals 6-9 is more opaque. Greenberg (1978: 291) suggests, with specific reference to Bantu, that such a "penumbra of the system" has to do with the lower frequency in use of these numerals compared to their lower counterparts. With this said, reflexes of these numerals exist far beyond the limits of the WS group, or even the GL branch. Arguably, some are reconstructible up to a putative Eastern Bantu ancestor (although arguing for such a proto-language is far beyond the scope of this article). At the same time, however, the different stems constituting this set of numerals in the WS language varieties also pattern differently in relation to genealogical and/or geographical parameters. One major feature which differentiates the WS members from some of their closest relatives/neighbours is the overall organization of this set of numerals. Thus, Jita (a Suguti language) and Gusii (classified within the North Mara group) make use of an additive quinary-based system of the form augend-link-addend, using the numerals 1-5 with 5 as a base, namely "5 and/with 1, 2, 3, 4…" (a structure similar to that used for forming 11-19 with a base-ten in WS, see subsection 2.1 above). The synchronic transparency of this system would suggest that it is a relatively more recent innovation (see

Schapper & Klamer 2014) which has possibly replaced cognates of the numerals found in the WS group. The fact that another Suguti language variety, Kwaya, has a system for forming 6-9 which is identical to that of WS would point towards such a conclusion. Furthermore, Gusii, unlike Jita, does not use the base-five approach to form '9'. Instead *kianda* is employed, that is, a reflex of the same form found in the WS group.

Another difference between the WS languages and their GL relatives, particularly outside the confines of the Mara branch, has simply to do with the lack of cognancy. This specifically concerns the words for '6' and '7', which are also the most problematic numerals to account for in terms of etymology. No obvious candidates emerge and the references that do discuss these forms do so in quite speculative, and occasionally contradictory manners.

Starting with the word for '6', *-saansaβa ~ -saasaβe*: similar to the formal division between Ngoreme and the rest of the WS members, reflexes of this form may be expressed with or without an additional word-medial /n/, e.g. Kuria *-sansaβa*, but Shashi *-sasaβa*. The WS language varieties pattern with most varieties of the Mara group in having this form of the numeral, with the exception of Gusii, which uses the five-based system mentioned above and hence lacks the form. However, except in the Mara language varieties, the spread of this form is limited, its cognates being confined to the very northern borders of (Eastern) Bantu and the varieties of the Greater Luhya (JE30) subgroup, such as (the clusters of varieties constituting) Masaba (JE31)[12] and Luhya (JE32). As these varieties neighbour the Mara group to the north without being considered to be directly related (cf. subsection 3.1), this shared numeral cognate is suggestive of an areal trait. In fact, Ehret (1971: 130) claims *sa(n)saβa* to be a loanword from Proto/Pre-South Nilotic **tɪsap* '7' (which he (1971: 111) ultimately links to a stem borrowed from Eastern Cushitic). However, although a plausible account of the phonological adaption involved is given (Proto-South Nilotic *t > s; *ɪ > *a*), the semantic motivation of meaning shift from '7' to '6' is not clear.

Johnston (1919-1922 vol. II: 469) instead suggests no less than three Bantu-inherent etymologies of *-sa(n)saβa*, of which none seems to work very well with the WS data. Two of these etymologies entertain the idea that *-sansaβa* is in

[12] Note that the sources cited for JE31 in a table in the Appendix both give (a reflex of) a Swahili borrowing for '7'. The deduction that JE31 languages have (had) forms cognate with *sa(n)saba* to express '7' is based on Johnston (1919-1922, vol. 1: 82) but also on the (unchecked!) data found in Chan's (1998-2019) online database on Numeral Systems of the World's Languages https://mpi-lingweb.shh.mpg.de/numeral/ [20.08.20].

some way derived from '3', in a fashion similar to °-*tanda*(*tu*), a common word for '6' in several Eastern Bantu languages (see Schmidl 1915, Meinhof 1948: 118, Hoffman 1953, Meeussen 1969). Although it is tempting to reconstruct *sa*(*n*)*saβa* in this way there are problems with such an endeavour. Specifically, one would have to explain why /t/ spirantized here and not elsewhere. Of course, as we saw in subsection 3.3, it may happen for (lexicalizing) numerals that phonemic change occurs sporadically without adherence to regular sound laws (see also Schapper & Klamer 2014) and the weakening of an already weak /t/ to a fricative would in that case not be a big leap. However, we have failed to find any evidence for this scenario. Moreover, we still would not know how to explain the ending /*βa*/.

The stem for '7', -*huŋgate*, has a different semasiological background from that of '6'. It is also attested for the whole of Mara (minus Gusii, plus Kwaya), and has a much wider distribution in the Eastern Bantu region. What is more, the most northern attestation of a reflex of this form in the GL group comes from Kuria (JE43), that is, a North Mara language and the closest neighbour to the WS members in the north. The Mara languages also seem to constitute the westernmost outpost for this particular numeral, with cognate forms attested only further to the east and the south-east, across much of Southern Kenya and Northern and Central Tanzania (see e.g. Werner 1919: 138, Hoffmann 1953: 171). It does not occur in the rest of the GL group, however, which makes it difficult to account for how this form entered the Mara branch. That is, is -*huŋgate* a retention from an early Eastern Bantu stem which disappeared in the rest of the GL or is it the result of diffusion from the east? As it surfaces in most of the Mara varieties, however, it is still fairly safe to conclude that it was inherited into WS branch.

Despite the relatively wide distribution of cognates of -*huŋgate* '7', the etymology is still opaque. Meeussen (1969: 17), citing Hoffmann (1953: 71-72), rejects Meinhof's (1948: 119) reconstruction (see also Schmidl 1915), which links the form to **-*túng*- 'tie' + *tatu* three. Instead, the (more fine-grained) comparative data demonstrate that the stem-initial consonant of the proto-form must have been **p* and not **t*. Such a reconstruction also makes perfect sense for the realization of the stem-initial consonant in WS, as **p* has regularly been debuccalized to /h/ in these language varieties. Meeussen's (1969: 19) own suggested etymon for this numeral is a verb °-*punk*- 'point, demonstrate' (and derivatives thereof), along the line of reasoning that the index finger would form the seventh finger when counting on the hands. However, as he points out himself, this is a very fragile reconstruction. Apart from some morphophonological problems

involved in connecting this verb stem with the numeral, a major difficulty is the fact that this verb stem is only attested in some Luba varieties (L30), and does not seem to occur in any Eastern Bantu languages, including in the WS group.

The word for '8' has often been suggested in the Bantu literature as being derived from the doubling of '4', from counting with four fingers on one hand and four fingers on the other (see, *inter alia*, Werner 1919: 134, Schmidl 1915, Meinhof 1948: 118, Greenberg 1978, Schadeberg 2003). Such a proposal also holds for *-naane* '8' in the WS language varieties. In fact, the WS data add valuable strength to such a proposal, as both of the nasals in this word form are palatalized when inflected with the agreement prefix of class 10, that is, *-iɲaaɲe* (see subsection 2.1). This would indicate that '8' to some extent is still analyzed as a composition of '4' and '4'.

The word for '9', *kénda*, can be linked to a stem **-kèndá* which is widespread across the East African part of Eastern Bantu, and particularly the Guthrie zones E, F, G, and J (= the GL), see Struck (1911: 991), Hoffmann (1953: 75), Guthrie (1967-1971 Vol III: 160, 275), and Bastin et al. (2002)[13]. Reflexes of this stem occur across the entire Mara group, including in the otherwise differing Gusii. Hence, it can be safely assumed that this numeral exists in the WS group through inheritance.

3.5. '10' and the decimal base(s)

The word *ikómi* used for '10' can be directly linked to a Proto-Bantu reconstruction, namely the stem *-kómì* 'ten' (noun class 5/6)[14] (e.g. Bastin et al 2002). Pozdniakov (2018: 133) traces this stem to an innovation **kum/kam/gham* in Bantoid, that is, a higher node in the (putative) Niger-Congo phylum of which the Bantu family is a part. Hence, it is probably even older than Proto-Bantu. The word used for forming multiple of tens, *murɔ́ŋgɔ* / *mirɔ́ŋgɔ*, has a long history as well. This lexeme is a reflex of the form **-dòngò* 'ten (decade)' (noun class 3/4), which is attested

---

[13] Reflexes of **-kèndá* are also attested in zone L (Bastin et al. 2002), belonging to (South-)Western Bantu (see Grollemund et al. 2015). There are also attestations of this numeral in Zone M of the "osculant pair" **yenda*, which additionally surfaces in some parts of the GL/zone J languages; see Guthrie (1967-1971, vol. III: 275; Bastin et al. 2002).

[14] The original noun class membership is most obvious in Ishenyi, which retains a reflex of the full prefix **di-* in ordinal connective constructions (see example (15) in subsection 2.2 above).

with a wide distribution within Eastern Bantu (see Guthrie 1967-1971 vol. III: 180 (=CS 663); Bastin et al. 2002)[15].

According to Guthrie, *mirɔ́ŋgɔ* arose to form numerals for multiples of '10' in Eastern Bantu, a role previously fulfilled by **-kɔ́mì*. This division of labour is in accordance with a general cross-linguistic tendency to have a "suppletive alternant" for the actual word for '10' to form decimals (Greenberg 1978). Accordingly, this pattern stretches through a large part of the Eastern Bantu area, including all Mara language varieties. However, it does not seem to be very widespread in the rest of the GL languages, which seem to prefer the use of pluralized reflexes of *-kɔ́mì* or other strategies where the base is not transparent[16]. Thus, the situation is similar to that described for the numeral '7' in subsection 3.4.

Notice that the additive link, used to form numerals within decades, is also reconstructed for Proto-Bantu as **nà* 'with, and' (Bastin et al. 2002). Meinhof (1948: 121) already notes its extended use across the Bantu family as a link when forming additive numerals. Forming additive numerals with a "comitative link" like this is also typologically common, as pointed out by Greenberg (1978). The ubiquitous spread of **na* with this function across the Bantu family strongly suggests that it was inherited into the ancestor of the WS language varieties.

In conclusion, it is safe to assume that both decimal stems were inherited into the WS. It is also likely that the division of labour between them was inherited, namely with *-komi* being used for the actual numeral 'ten' and as the base for forming additive numerals (in a construction with a comitative link) and with *-roŋgo* as a dedicated base for multiplication.

3.6. The higher base numbers (100, 1000, 100 000)

The word for '100', *riiɣána*, in the WS language varieties can be linked straightforwardly with a stem **-gànà* (noun class 5/6), which, according to Guthrie (1967-71, vol. III: 206 (=CS 774)) is "probably" a Proto-Eastern Bantu item given its distribution throughout zones D to zone S (see also Dempwolff 1916-1917: 137; Bastin et al. 2002). Indeed, many other GL languages surveyed for this

---

[15] Bastin et al. (2002) further associates **dòngò* with the nominal stem **-dòngò* 'line, row' and, ultimately, the verb **-dóng-* 'heap up', both with a wide distribution throughout the Bantu speaking area, thus arguably being reconstructible even for Proto-Bantu.

[16] One exception is Ha (JD66), a Western Lakes language remotely related to Mara, where forming decimals up to 50 can be done freely with either a reflex of *kɔ́mì* or of **dòngò* (Harjula 2004: 78-79).

study, including all Mara languages, have a reflex of this stem, even Gusii[17]. Thus, it is definitely a shared retention among the WS language varieties.

Reflexes for the word for '1000', *eɣe-kwé* (Ikoma *eɣe-kú* due to the restriction on labialization on the final syllable), is not attested in any general Bantu reconstruction work. However, there are scattered attestations across the Eastern Bantu area. It has been reconstructed for Proto-Sabaki (~G40, E70) by Nurse & Hinnebusch (1993: 292, 663), see also Nicolle (2013: 39-40) on Digo (E73) specifically. It is further attested in Chewa-Nyanja (N31; Werner 1919: 140) and Cuwabo (P34; Guérois 2019). Except for Gusii (which only has a Swahili borrowing attested), all members of the Mara branch – and some of the Suguti – use a cognate of this stem for '1000'. Otherwise, *eɣe-kwé*, and reflexes thereof, is once again a numeral that does not seem to be used in other parts of the GL area.

Any source meaning of the stem is not clear. No further meanings were provided by the consultants. Is it somehow connected to the proto-GL stem **-kwé* (noun class 14) 'bride price' (see Schoenbrun 1997: 94-95) and/or the stem *-kwe ~ -ku* in the WS language varieties, meaning 'firewood' (as in a heap or a pile, a common type of metaphorical extension for large numerals in Bantu, see Schmidl 1915)?

Similarly, *-rara* (noun class 7/8) for '100 000' remains an enigma from an etymological point of view, not least because it is seldom the case that a digit of such high value is mentioned in the comparative literature. However, as this numeral surfaces in all the WS language varieties it is at least a shared retention or innovation within this group.

3.7. Ordinals

As already mentioned in subsection 2.2, the connective construction for forming ordinals is common across the Bantu speaking area and thus is also undoubtedly an inherited pattern in the WS group. To that we may add Polak-Bynon's (1965:

---

[17] However, the case of Gusii is not as clear-cut as for the other members of the Mara group. Whiteley (1965: 18, also Cammenga 2002: 356) gives *e-mia* for '100', treating *ri-gana* as an alternative form. Whiteley (1965: 18) additionally suggests *ri-gana* as a borrowing from Luo (= Dholou, Western Nilotic). Luo does seem to have a stem gana for '1000' (see Tucker 1994: 265). However, given the abundance of reflexes of this stem in the Bantu languages, both within and outside the geographical reach of contact with Luo, it would be more reasonable to believe that the diffusion has gone the other way around. The word *mia*, however, was most likely borrowed into Gusii from Swahili (which in turn copied it from Arabic *miʾa(t)* 'hundred'; see Schadeberg 2009).

136) statement that connective ordinals accompanied by the preceding element *ka-* are particularly common in (North)Eastern Bantu in general and particularly around Lake Victoria (i.e. GL). Polak-Bynon (1965: 136) links this element to the noun class 12 prefix, which has the same form, and which is also commonly used for deriving adverbials from nominal stems (e.g. Meinhof 1948: 124).

Regarding the deviant derivatives of 'first', we may note the following. To begin with, *-mbɛrɛ*, used exclusively for 'first' in Ikoma, Nata and Ishenyi, and variably used in Ngoreme, originates from a form meaning '(in) front (of)', as seen in (37) below. (Similarly, *ɲuma* 'back' is exclusively used to denote 'last' in Ikoma, Nata and Ishenyi, as well as in Ngoreme.)

(37) Ikoma

a-ŋómbe	*n-e-eɲí*	*á-mbɛrɛ*	*e*	*mé-te*	
9-cow		FOC-SP9-COP	9-front	CONN9	3-tree

'The cow is in front of the tree.'

To derive the ordinal numeral 'first' from 'front' is an extremely common pattern across Eastern Bantu, and languages using this strategy include members from most subgroups of GL, although its adjectival use with nominal prefixes rather than within a connective construction is innovative (see Polak-Bynon 1965: 150-151, see also Grégoire 1975: 212-215). Taken together with the fact that *mbɛrɛ* for 'first' is attested for all WS language varieties it is most likely a shared retention.

The irregular ordinal *kwánsa* in Ngoreme is probably a borrowing from Swahili, partly given the variation from *mbɛrɛ* but also because a reflex of the source verb **-yànd-* 'begin' (Nurse & Hinnebusch 1993: 650) does not seem to exist in the language. Polak-Bynon (1965: 149) finds related forms in other Eastern Bantu varieties such as Pogoro and Sukuma, but mentions that these are likely to be borrowings from Swahili as well. See also Greenberg (1978), who notes that unlike the ordinal '1', the borrowing of the equivalent cardinal 'first' is not uncommon cross-linguistically.

4. Summary and conclusions

In this study we have accounted for the synchronic and diachronic aspects of the numeral system in four closely related Bantu varieties of the Western Serengeti subgroup, that is, Ikoma, Nata, Ishenyi, and Ngoreme. The study has shown that the systems are more or less identical. Except for the borrowed ordinal *kwánsa* in Ngoreme, the numerals only differ in terms of the more overarching

phonological disparities characterizing the different members of this group.

We can furthermore conclude that the WS language varieties are conservative with regard to this specific linguistic domain, particularly the fact that they have kept the conditioned weakening of the numerals '3'-'5' with the agreement prefix of class 10 suggested for Proto-Bantu. It is possible that this maintenance is connected to the fact that "Bantu spirantization" (Bostoen 2008) has generally not affected these language varieties. However, most other simplex numerals and strategies of forming complex numerals can also be linked to cognates with a further distribution across the (Eastern) Bantu family. This stability in itself can be taken to stand out, not least as other parts of these language varieties are remarkably different from canonical Bantu patterns (such as a highly complex vowel harmony system, inverted auxiliary constructions, and *non*-inverted existentials).

The fact that the numeral systems of the WS language varieties (and the Mara language varieties more generally) appear to pattern more closely with other Eastern Bantu languages than with their supposedly closest relatives of the GL group is problematic from a genealogical perspective and would bring support to Nurse's (1999) scepticism with regard to this grouping. In addition to this, the etymologies of some of the numerals are still left unresolved. These two facts, taken together, serve as an impetus for further comparative work on this subject and in this region.

Finally, we note that the only clearly attested Swahili borrowing in the language varieties is (most likely) *kwánsa* 'first' in Ngoreme. This is in contrast with other Tanzanian Bantu languages where especially the higher numerals are claimed to have shifted to Swahili to a more or lesser degree (see e.g. Morrison 2011: 216, Bernander 2017: 32, 79, Wilhelmsen 2019). Hence, it would seem that the WS numeral systems are not in danger of extinction, or at least not at the moment.

5. Abbreviations and symbols

* = Reconstructed form

° = Tentatively reconstructed form

Numbers refer to noun classes

AUG	augment
CAUS	causative
CONN	connective

COP	copula
DEM	demonstrative
DIST	distal
FOC	focus
FV	final vowel
IPFV	imperfective
LOC	locative
OP	object prefix
PFV	perfective
POSS	possessive
PROG	progressive
PROX	proximal
PST	past
REF	referential
SBJV	subjunctive
SG	singular
SP	subject prefix

References

Appleby, L. L. 1961. *A first Luyia grammar with exercises*. Nairobi: East African Literature Bureau.

Aunio, L., H. Robinson, T. Roth, O. Stegen & J. B. Walker. 2019. "The Mara languages (JE40)". *The Bantu languages*, 2nd edn., ed. by K. Bostoen, D. Nurse, G. Philippson & M. Van de Velde. London: Routledge. 501-532.

Austen, C. L. 1974. *Aspects of Bukusu syntax and phonology*. Unpublished PhD thesis. Indiana University.

Bastin, Y., A. Coupez, E. Mumba, & T. Schadeberg (eds). 2002. *Bantu lexical reconstructions 3 / Reconstructions lexicales bantoues 3*. Tervuren: Royal Museum for Central Africa.

Bernander, R. 2017. *Grammar and grammaticalization in Manda: An analysis of the wider TAM domain in a Tanzanian Bantu language*. Unpublished PhD thesis. Gothenburg University.

Blažek, V. 1999. *Numerals: Comparative-etymological analyses of numerical systems and their implications*. Masaryk University, Brno.

Bostoen, K. 2008. "Bantu spirantization, morphologization, lexicalization and historical classification". *Diachronica* 25(3). 299-356.

Cammenga, J. 2002. *Phonology and morphology of Ekegusii: A Bantu language of Kenya*. (East African Languages and Dialects 12.) Cologne: Rüdiger Köppe.

Cammenga, J. 2004. *Igikuria phonology and morphology: A Bantu language of South-West Kenya and North-West Tanzania*. Cologne: Rüdiger Köppe.

Comrie, B. 2005a. "Endangered numeral systems". *Bedrohte Vielfalt: Aspekte des Sprach(en)tods*, ed. by J. Wohlgemuth & T. Dirksmeyer. Berlin: Weißensee. 203-230.

Comrie, B. 2005b. "Numeral bases". *World atlas of language structures*, ed. by M. Haspelmath, M. S. Dryer, B. Comrie & D. Gil. Oxford: Oxford University Press. 530-533.

Crabtree, W. A. 1923. *Elements of Luganda grammar, together with exercises and vocabulary*. Kampala: Uganda Bookshop. Society for Promoting Christian Knowledge.

Davy, J. & D. Nurse. 1982. "Synchronic versions of Dahl's law: The multiple applications of a phonological dissimilation rule". *Journal of African Languages and Linguistics* 4. 157-195.

De Blois, K. 1970. "The augment in the Bantu languages". *Africana Linguistica* 4. 85-165.

Dempwolff, O. 1914-1915. "Beiträge zur Kenntnis der Sprachen in Deutsch-Ostafrika 4: Kulia". *Zeitsschrift für Kolonialsprachen* 5. 26-44. 113-136.

Dempwolff, O. 1916-1917. "Beiträge zur Kenntnis der Sprachen in Deutsch-Ostafrika 9: Ostbantu-Wortstämme". *Zeitschrift für Kolonialsprachen* 7. 134-160. 167-192.

Dimmendaal, G. "Language shift and morphological convergence in the Nilotic area". *Sprache und Geschichte in Afrika* 16-17. 83-124.

Ehret, C. 1971. *Southern Nilotic history: Linguistic approaches to the study of the past*. Evanston: Northwestern University Press.

Futakamba, W., S. Koren, H. Robinson & O. Stegen. 2013. *The Grammar Basics of Zanaki*. Unpublished Paper: SIL International.

Gil, D. 2013. "Distributive Numerals". *World atlas of language structures online*, ed. by M. Haspelmath, M. S. Dryer, B. Comrie & D. Gil. Oxford: Oxford University Press. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: http://wals.info/chapter/54 [6.07.2020].

Greenberg, J. H. 1978. "Numeral Systems". *Universals of human language* 3: *Word structure*, ed. by J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik. Stanford: Stanford University Press. 250-295.

Greenberg, J. H., 2000. "Numeral". *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung / Morphology: An international handbook on inflection and word formation* 1, ed. by G. Boij, C. Lehmann & J. Mugdan. Berlin: Mouton de Gruyter. 770-782.

Grégoire, C. 1975. *Les locatifs en bantou* (Annales Sciences Humaines 83). Tervuren: Musée Royal de l'Afrique Centrale.

Grimm, N. 2019. "Review of Pozdniakov, K., The numeral system of Proto-Niger-Congo: A step-by-step reconstruction". *Linguistique et Langues Africaines* 5. 115-122.

Grollemund, R., S. Branford, K. Bostoen, A. Meade, C. Venditti & M. Pagel. 2015. "Bantu expansion shows that habitat alters the route and pace of human dispersals". *Proceedings of the National Academy of Sciences of the United States of America* 112 (43). 13296-13301.

Guérois, R. 2019. "Cuwabo P34". *The Bantu Languages*, 2nd edn., ed. by K. Bostoen, D. Nurse, G. Philippson & M. Van de Velde. London: Routledge. 733-775.

Guthrie, M. 1948. *The classification of the Bantu languages*. London: Oxford University Press.

Guthrie, M. 1967-1971. *Comparative Bantu: An introduction to comparative linguistics and prehistory of the Bantu languages*. Vol. 1-4. Farnborough: Gregg International.

Güldemann, T. 2018. "Historical linguistics and genealogical language classification in Africa". *The Languages and Linguistics of Africa*. (The World of Linguistic 11), ed. T. Güldemann. Berlin: Mouton de Gruyter. 58-444.

Hammarström, H. 2010. "Rarities in numeral systems". *Rethinking universals: How rarities affect linguistic theory*, ed. by J. Wohlgemuth & M. Cysouw. Berlin: Mouton de Gruyter. 11-60.

Haji, S. 2000. *Haya vocabulary* (Asian and African lexicon series 37). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Harjula, L. 2004. *The Ha language of Tanzania*. Cologne: Rüdiger Köppe.

Hill, D., A.-L. Lindfors, L. Nagler, M. Woodward & R. Yalonde. 2007. *A sociolinguistic survey of the Bantu languages in Mara region*, Tanzania. Unpublished Manuscript. SIL International.

Hoffmann, C. 1953. "Zur Verbreitung der Zahlwortstämme in Bantusprachen". *Afrika und Ubersee* 37. 67-80.

Hurford, J. R., 1987. *Language and number.* Oxford: Basil Blackwell.

Johnston, H. H. 1919-1922. *A comparative study of the Bantu and semi-Bantu languages*. Oxford: Clarendon Press.

Johnston, H. H. 1920. "The Ki-Shashi (Ki-şaşi) Language (S. E. Victoria Nyanza)". *Journal of the Royal African Society.* 19. 210-213.

Kagaya, R. 2005. *A Jita vocabulary*. (Asian and African Lexicon 47). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.

Kihore, Y. M. 2000. "Historical and linguistic aspects of Kihacha". *Lugha za Tanzania / Languages of Tanzania: Studies dedicated to the memory of Prof. Clement Maganga*, ed. by K. K. Kahigi, Y. M. Kihore & M. Mous. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), State University of Leiden. 67-80.

Laine, A. O. 2016. *Isenjen kielen vokaalijärjestelmä* [*The vocal system of the Isenje language*]. Unpublished MA thesis. University of Helsinki.

Legère, K. 2006. "Language endangerment in Tanzania: Identifying and maintaining endangered languages". *South African Journal of African Languages* 26. 99-112.

Maho, J. F. 1999. *A comparative study of Bantu noun classes*. Göteborg: Acta Universitatis Gothoburgensis.

Maho, J. F. 2003. "A classification of the Bantu languages – an update of Guthrie's referential system". *The Bantu Languages*, ed. by D. Nurse & G. Philippson. New York: Routledge. 639-651.

Maho, J. F. 2009. *NUGL Online: The Online Version of the New Updated Guthrie List, a Referential Classification of the Bantu Languages*. Unpublished Paper. Online: https://pdfs.semanticscholar.org/887a/3570883e1810b2c81be86d139617dfd89b0c.pdf [28.06.2020].

Meeussen, A. E. 1967. "Bantu grammatical reconstructions". *Africana Linguistica* 3. 79-121.

Meeussen, A. E. 1969. "Bemerkung über die Zahlwörter von sechs bis zehn in Bantusprachen". *Kalima na dini: Studien zur Afrikanistik, Missionswissenschaft, Religioswissenschaft; Ernst Dammann zum 65. Geburtstag*, ed. by H.-J. Greschat & H. Jungraithmayr. Stuttgart: Evangelische Missionsverlag. 11-18.

Meinhof, C. 1948. *Grundzüge einer vergleichenden Grammatik der Bantusprachen*, 2nd rev. edn. Hamburg: Eckardt & Messtorff.

Morrison, M. 2011. *A reference grammar of Bena*. Unpublished PhD thesis. Rice University.

Mreta, A. Y. 2008. *Kisimbiti: Msamiati wa Kisimbiti-Kiingereza-Kiswahili na Kiingereza-Kisimbiti-Kiswahili / Simbiti-English-Swahili and English-Simbiti-Swahili Lexicon*. (LOT Publications Lexicon Series 7). Dept. of Foreign Languages and Linguistics, University of Dar es Salaam.

Mutonyi, N. 2000. *Aspects of Bukusu morphology and phonology*. Unpublished PhD thesis. Ohio State University.

Nurse, D. 1999. "Towards a historical classification of East African Bantu Languages". *Bantu historical linguistics: Theoretical and empirical perspectives*, ed. by J.-M. Hombert & L. Hyman. Stanford: CSLI. 1-41.

Nurse, D. & G. Philippson. 1980. "The Bantu languages of East Africa: A lexicostatistical survey". *Language in Tanzania*, ed. by E. C. Polomé & C. P. Hill. Oxford: Oxford University Press. 26-67.

Odom, S. R. 2016. *The Grammar Basics of Kwaya*. Unpublished Paper: SIL International. Online: https://www.sil.org/resources/archives/67534 [27.06.2020].

Odom, S. R. & H. Robinson. 2016. *The Grammar Basics of Jita*. Unpublished Paper: SIL International. Online: https://www.sil.org/resources/archives/67533 [27.06.2020].

Overton, R. & J. B. Walker. 2017. *The Grammar basics of Kabwa*. Unpublished Paper: SIL International. Online: https://www.sil.org/resources/archives/73083 [27.06.2020].

Philippson, G. & R. Grollemund. 2019. "Classifying Bantu languages". *The Bantu Languages*, 2nd edn., ed. by K. Bostoen, D. Nurse, G. Philippson & M. Van den Velde. London: Routledge. 335-354.

Polak-Bynon, L. 1965. "L'expression des ordinaux dans les langues bantoues". *Africana Linguistica* 2. 127-159.

Polak-Bynon, L. 1975. *A Shi grammar: Surface structures and generative phonology of a Bantu language*. (Annalen van het Koninglijk Museum van Belgisch-Congo: Reeks in 8, 86). Tervuren.

Pozdniakov, K. 2018. *The numeral system of Proto-Niger-Congo: A step-by-step reconstruction*. Berlin: Language Science Press.

Rankin, R. 2006. "The interplay of synchronic and diachronic discovery in Siouan grammar–writing". Catching Language: *The standing Challenge of Grammar Writing*, ed. by F. Ameka, A. Dench & N. Evans. Berlin, Boston: De Gruyter Mouton. 527-547.

Rischel, J. 1997. "Typology and reconstruction of numeral systems: The case of Austroasiatic". *Linguistic Reconstruction and Typology*, ed. by J. Fisiak. Berlin: Mouton de Gruyter. 273-312.

Robinson, H. & M. Sandeen. 2015. *The grammar basics of Kabwa*. Unpublished Paper: SIL International. Online: https://www.sil.org/resources/archives/73083 [27.06.2020].

Roth, T. 2018. *Aspect in Ikoma and Ngoreme: A comparison and analysis of two Western Serengeti Bantu languages*. Unpublished PhD dissertation. University of Helsinki.

Roth, T. & H. Gibson. 2019. "Differential diagnoses of language change in Ngoreme, a Bantu Language of Tanzania". *Africana Linguistica* 25. 281-319.

Schadeberg, T. 2003. "Historical linguistics". *The Bantu languages*, ed. by D. Nurse & G. Philippson. New York: Routledge. 143-163.

Schadeberg, T. 2009. "Loan words in Swahili". *Loanwords in the world's languages: A comparative handbook*, ed. by M. Haspelmath & U. Tadmor. Berlin: De Gruyter Mouton. 76-102.

Schapper, A. & M. Klamer. 2014. "Numeral systems in the Alor-Pantar languages". *The Alor-Pantar languages: History and typology*, ed. by M. Klamer. Berlin: Language Science Press. 285-336.

Schmidl, M. 1915. "Zahl und Zählen in Afrika". *Mitteilungen der Anthropologischen Gesellschaft in Wien* 45. 165-209.

Schoenbrun, D. L. 1990. *Early history in Eastern Africa's Great Lakes region: Linguistic, ecological and archaeological approaches, ca. 500 B.C. to ca. 1000 A.D.* Dissertation. UCLA, Los Angeles.

Schoenbrun D. L. 1994. "Great Lakes Bantu: Classification and settlement chronology". *Sprache und Geschichte in Afrika* 16: 91-152.

Schoenbrun, D. L. 1997. *The historical reconstruction of Great Lakes Bantu cultural vocabulary: Etymologies and distributions*. Cologne: Rüdiger Köppe.

Shetler, J. B. 2003. *Telling our own stories: Local histories from South Mara, Tanzania*. (African Sources for African History 4). Leiden: Brill.

Sillery, A. 1932. "A sketch of Kikwaya grammar". *Bantu Studies* 6. 273-308.

Sillery, A. 1936. "Notes for a grammar of the Kuria language". *Bantu Studies* 10. 9-30.

Stappers, L. 1965. "Het hoofdtelwoord in de Bantoe-talen". *Africana Linguistica* 2. 177-199.

Stolz, T. & L. Veselinova. 2013. "Derivation and suppletion of ordinal numerals". *World atlas of language structures online*, ed. by M. Haspelmath, M. S. Dryer, B. Comrie & D. Gil. Oxford: Oxford University Press. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Struck, B. 1911. "Die Fipasprache". *Anthropos* 6. 951-994.

Thornell, C. 2004. "The Noun Phrase in the Kerebe Language". *Globalisation and African languages: Risks and benefits. Festschrift Karsten Legère*, ed. by K. Bromber & B. Smieja. Berlin & New York: Berlin: Mouton de Gruyter. 219-242.

Tucker, A. N. 1994. *A grammar of Kenya Luo (Dholuo)*. (Nilo-Saharan: Linguistics Analyses and Documentation 8). Cologne: Rüdiger Köppe.

Valinande, N. K. 1984. *The structure of Kinande*. Vol. 1-4. Unpublished PhD thesis. Georgetown University.

Van de Velde, M. 2013. "The Bantu connective construction". *The Genitive*, ed. by A. Carlier & J.-C. Verstraete. Amsterdam & Philadelphia: John Benjamins. 217-252.

Van de Velde, M. 2019. "Nominal morphology and syntax". *The Bantu Languages*, 2nd rev. edn., ed. by K. Bostoen, D. Nurse, G. Philippson & M. Van de Velde. London: Routledge. 237-269.

Vanhoudt, B. 1994. "L'expression de 'un' dans la numération référentielle en protobantou". *Africana Linguistica* 10. 215-221.

Walker, J. & R. Overton 2018. *The grammar basics of Simbiti*. Unpublished Paper: SIL International.

Werner, A. 1919. *Introductory sketch of the Bantu languages*. London: Trübner.

Whiteley, W. H. 1965. *A practical introduction to Gusii*. Dar es Salaam: The East African Literature Bureau.

Wilhelmsen, V. 2019. "Mbugwe F34". *The Bantu Languages*, 2nd rev. edn., ed. by K. Bostoen, D. Nurse, G. Philippson & M. Van de Velde. London: Routledge. 533-562.

Zerbian, S. & M. Krifka. 2008. "Quantification across Bantu languages". *Quantification: A cross-linguistic perspective*. (North-Holland Linguistic Series: Linguistic Variations Volume 64), ed. by L. Matthewson. Bingley, UK: Emerald. 383-414.

Appendix

Great Lakes. Sources surveyed for comparative data

Guthrie code	Language	References
JD42	Nande	Valinande (1984: 653-708)
JD53	Shi	Polak-Bynon (1975: 414-421)
JD66	Ha	Harjula (2004: 77-81)
JE15	Ganda	Crabtree (1923: 165)
JE22	Haya	Kaji (2000: 101-111)
JE24	Kerebe	Thornell (2004: 237-238)
JE25	Jita	Kagaya (2005: 454-455), Odom & Robinson (2016: 9-10)
JE251	Kwaya	Odom (2016: 9-10), Sillery (1932: 276-277)
JE31c	Bukusu (Masaba)	Austen (1974: 131-135), Mutonyi (2000: 105-109)
JE32	Lu(h)yia	Appleby (1961: 19-21)
		MARA SUBGROUP
JE402	Ikizu	Robinson & Sandeen (2015: 10-11)
JE404	Shashi/Sizaki	Johnston (1920: 212)
JE405	Kabwa	Overton & Walker (2017: 11)
JE42	Gusii	Whiteley (1965: 17), Cammenga (2002: 348-367)
JE43	Kuria	Sillery (1936: 14-15), Dempwolff (1914-1915: 122)
JE431	Simbiti	Mreta (2008: passim), Walker & Overton (2018: 11-12)
JE432	Hacha	Kihore (2000: 68)
JE44	Zanaki	Futakamba et al. (2013: 4), Hill et al. (2007)

A large number of numeral systems of other (Eastern) Bantu and Nilotic languages have also been checked, as they occur in the comparative literature. The data from Chan's (1998-2019) online database of Numeral Systems of the World's Languages https://mpi-lingweb.shh.mpg.de/numeral/ has also been considered; however, no data is cited from this website as it is explicitly made clear that the data need further checking for typos and errors (also E. Chan pers. comm. 18 Feb. 2020).