Nordic Journal of African Studies 8(2): 108-138 (1999)

NILO-SAHARAN REVISITED

PERTTI MIKKOLA
University of Helsinki, Finland

INTRODUCTION

Nilo-Saharan is the most problematic case in Greenberg's (1963/1966) classification of African languages. Previously the role of chance in mass comparison has been investigated with the *n/m*-ary approach (Mikkola 1998). The test comprised a standard sample from 18 languages in Greenberg's Nilo-Saharan word lists. Consonants were grouped into 8 types corresponding to natural classes, any vowels were accepted. The probabilities of every different word-initial CVC-sequence type were investigated, and the expected distributions were compared with the observed scores. Around a dozen of the observed similarities were more common than expected by random coincidence, though usually not significantly. For details of the methodology, results and the background see Mikkola (1998).

The aim of this paper is to go beyond the results of that statistical approach. Those of Greenberg's etymologies observed more often than expected are included in this more detailed qualitative investigation. These best etymologies are compared with new and more comprehensive data, including all Nilo-Saharan stocks. Besides the statistically strict tokens, other important supporting (and conflicting) evidence is taken into consideration. In addition, the results are controlled against other African phyla, including a short discussion concerning the Kongo-Saharan or Niger-Saharan hypotheses.[1]

---

[1] Abbreviations: Bender's (e.g. 1996a) code system for Nilo-Saharan (A, B, ... Ek, En /E1, E2, ...Fp, Fc ..., K, L) was applied; otherwise:

NS = Nilo-Saharan
CN = Chari-Nile
ES = Eastern Sudanic
NC = Niger-Congo
AA = Afroasiatic
KS = Khoisan
PWS = Proto-Western Sudanic (by Westermann 1927)
PHS = Proto-Hamito-Semitic (by Orel & Stolbova 1995)

CVC-types (Consonant-Vowel-Consonant), for details see Mikkola (1998: 71-72):
Ø = zero consonant (word-initial/final vowel/semi-vowel), M, N, L (l, r), S, P, T, and K.

1. SUPPORTING EVIDENCE

1.1 STRICT CRITERIA, GREENBERG'S DATA

The total length of the test list[2] was 93 items; altogether 68 glosses involving ca. 103 etymologies were observed to occur in Greenberg's Nilo-Saharan etymology lists. In the statistical approach only one of these (#225[3] in the combined list) 'who?' was significantly above the expected by random coincidence (at the 95 per cent level). In addition, nearly significant cases were etymologies #209 'tooth' (in Eastern Sudanic) and #224 'white 3'. Other important etymologies within the sample found more often than expected included notably: 'meat 2' (one of the common African roots), 'dog' (which is a Wanderwort), 'mouth 2' (Eastern Sudanic), 'man 1' (mostly Eastern Sudanic) and 'to kill 2/to die 2' (two roots; also found in Niger-Congo).

Though the quantity of strict similarities between Nilo-Saharan lineages is usually insignificant, we ought not to forget that some languages had incomplete data. None of the investigated etymologies occurred in more than half the sample languages (with strict criteria), in the original data.

1.2 SUPPLEMENTARY EVIDENCE

In search of additional supporting evidence, Bender (various works, especially 1996a), Ehret (1989 etc.), and several other sources were consulted. Gumuz and Krongo were also included in this search. The supplementary evidence will be discussed from linguistic and areal viewpoints. If statistically strict criteria are required, important additional supporting evidence was observed only in a few cases. Allowing transitional correspondences, more sound changes, semantic shifts, and coincidences in additional languages within the stocks, the number of observed possible cognates markedly increased. However, the role of chance increases considerably.

Nevertheless, the results seem to lend support to the Nilo-Saharan hypothesis, with the exception of only a few lineages. With supplementary evidence, one of Greenberg's proposed etymologies was observed to occur in 12 sample languages (out of 18). The expanded database also revealed further supporting evidence for other proposed cognates. For details see chapter 4. Because the statistical analysis was only based on a

---

[2] Bender's Ethiopian word list; without personal pronouns, which Greenberg included in his list of grammatical elements.

[3] Note that the numbering system was made for the combined word list; regarding the original (Greenberg 1963/1966) numbers see the discussion in chapter 4.

sample data of Greenberg's word lists, a few additional coincidences (i.e. other glosses) could presumably be found if comprehensive data were used for the whole list.

2. Coincidences in Individual Sample Languages

Using strict criteria (see Mikkola 1998), the individual sample languages featured these etymologies quite sporadically as seen in Appendix 2 (being more common than expected according to the Poisson distribution, though not necessarily significantly). Some chance cognates, loans (like Wanderwort 'dog') and possibly sound symbolic words are necessarily included in the table. Temein had no strict non-chance word-initial CVC-coincidence in the original sample, but this was due to the poor data. Songai, Bari, Mangbetu, and Koma had only 1 each. The most frequent sample languages were Maba and Nera, both with 6 items in the test list. All others featured 2-4 strict non-chance CVC-coincidences in the list.

Concerning the strict similarities, the supplementary data did not affect these figures significantly in most cases. Noteworthy were a few additional items in Eastern Sudanic (like Temein, Nyimang and Gaam) and a reduction of one or two items in Nyangi (not being strict coincidences).

If transitional coincidences (like 'movable k' and some differences in the second consonant) are included, the picture substantially changes. From several languages, considerably more evidence can be found. However, some languages, especially Songai and Nyangi, are not much better Nilo-Saharan candidates after this procedure. Temein and Bari are still badly represented. However, additional evidence was observed in other Temein and Nilotic languages. All other Nilo-Saharan sample languages featured at least 3-6 coincidences with this somewhat subjective approach (some Eastern Sudanic languages even 7-9). When moderate semantic shifts and cognates in closely related languages were allowed, still more evidence was observed - up to 10-11 roots per stock (out of 13 etymologies in Appendix 1).

In the non-sample Nilo-Saharan lineages, several important coincidences were found in Krongo, but only one or two in Gumuz. No definite solution for this can as yet be suggested. It might be that Gumuz is an isolate, instead of being a Nilo-Saharan language (Bender, p.c.). Note, that Bender (1979: 40) had even formerly expressed similar doubts: "perhaps not even a Sahelian language at all".

3. Is Nilo-Saharan both an inclusively and an exclusively valid lineage?

The comparison revealed several similarities between Nilo-Saharan and Niger-Congo requiring linguistic explanation. The similarities with Niger-Congo are often systematic.

In addition, they were found to occur in many different branches without even contact to Nilo-Saharan. Therefore, at least a late contact is, presumably, out of consideration as an explanation for the coincidences. Interestingly, this might support the Kongo-Saharan hypothesis proposed by Gregersen (1972).

Blench (1995) has claimed that Niger-Congo is only a 'branch' of Nilo-Saharan. Though his evidence for 'Niger-Saharan' is not exhaustive, the hypothesis might still be regarded as a serious one; however, possibly without some outliers, like Songai, Kuliak, and Gumuz. Nevertheless, I cannot agree with the details of his classification. See also the discussion and additional suggested items linking NS and NC by Bender (1981: 263; 1992: 37; and 1996a: esp. 66, 118-119, 126-136). Cf. also Boyd (1978, 1996) and Williamson (1989: 7-9). However, not all proposed similarities seem acceptable.

The few coincidences with Afroasiatic are presumably due to contact, sound symbolism, and chance. These similarities were generally observed to occur only in some limited areal contexts (especially in Chadic, Cushitic and Omotic), where borrowing is the most plausible explanation. South African Khoisan (and several non-African languages) appeared to be much more divergent, reflecting geographic distance and the lack of genetic relationship. As a whole, a handful of similarities is easy to find, even in unrelated languages (e.g. Sandawe).

Because the starting point in the statistical analysis was Nilo-Saharan, some bias follows and might distort the results regarding outside comparisons, i.e. towards a too comprehensive picture of Nilo-Saharan. Any consideration of this type of bias has often been neglected in language comparisons.

4. Comments on The Best Etymologies

No systematic error-searching concerning the quality of Greenberg's data was carried out, but note some (only minor) differences to other sources. Besides Greenberg's word lists, several other works were also consulted.[4] These are quoted only if additional supporting or conflicting evidence was found. Therefore, the lack of citation is relevant, and reflects that unrelated roots were observed to occur in these Nilo-Saharan stocks (in languages investigated). To save space, the differences in transcriptions and the (often exhaustive) additional evidence within the same stocks are generally not shown (unless qualitatively relevant). The etymologies are discussed in alphabetical order, in each case beginning with the original data in Greenberg (1966). It has to be emphasized that this catalogue-type discussion includes, besides real cognates, also random coincidences and loan words.

[4] I have to regret the unsatisfactory quality of some Omotic data.

The etymology in the combined list	Word-initial CVC-type	The etymologies in the original word lists of Greenberg (1966)
e#60 'dog'	**ØVS**	**(NS#46, CN#31, ES#36)**

Similar word-initial CVC-sequence occurred in 4 sample languages, only 2 were expected by random coincidence.

D	Fur	*asa*	
E3	Nera	*wos*	
E7	Merarit	*wi:s*	
E8	Daju	*i:si*	(Daju='Dagu of Darfur', unless otherwise stated)

In addition, 2 'dissimilar' words in the sample languages:

A	Gao	*hanši*;	Cf. 'hound'
F	Mangbetu	*si*,	Mittu *wisi*

Note also the following forms, originally from S[outhern] C[ushitic] according to Ehret, as cited by Fleming (1983: 439):

K	Tepes (Soo)	*suyan* 'dog, wild' (*Lycaon Brooks*)
E9	Maasai	*o-suyiani*

One of Bender's (1996a: 143) 'Nilo-Saharan fragments'. See also Bender (1981: 258). According to Blench (1995: 127), the root *#-si* 'dog' is "extremely widespread in Central Africa", e.g.

C	Masalit	*wasi*;	*ínji* in Edgar (1991b: 126)
F	Baka	*ísì*	
L	Katcha	*ìs(s)ì*;	Tolibi (Katcha) *ʈeerá/(ʈ̪)ʈ̪ɪnɪ́* (Schadeberg 1994: 26)
NC	Nupe	*efi*	

The root is more common than expected due to chance. However, this is a Wanderwort having corresponding forms in unrelated languages: including Ethiosemitic, e.g. Amharic and Gurage; and Highland Cushitic, e.g. Sidamo and Kembata (see e.g. Bender 1971). Presumably of Afroasiatic origin: Orel & Stolbova's (1995: 536) PHS #2571 **yaĉ-/*wa-yaĉ-* 'dog':

AA[5]	Ometo	*wayše*	
	Highland East Cushitic	**wis-*	
	Beja	*yaas*	
	Egyptian	*iš* pl. 'dogs pulling the ship of the Sun-god'; and 'derivative' in	
	Berber	**wVs[i]n* 'jackal'	

e#91 'to go 1'	**LVØ**	**(CN#43, ES#55)**

Word-initial LVØ-coincidence was found in 2 sample languages; only 1 occurrence was expected:

E7	Merarit	*la*	
H	Kunama	*li*;	*'lau-* in Bender (1996a: 101); *ga:s-* (Bender 1971: 277)

Note the following forms outside the sample:

E9	Dinka	*lɔ*, Maasai *lo*
F	Lugbara	*lU*, Kreish *lo:*

Note also (possible supplementary evidence):

B	Kanuri	*lengîn, lejîn* (Cyffer 1994: 80)
E3	Nera	*ɛ:l-* (Bender 1971: 268)

Cf. Bender's (1996a: 101) "Fair isogloss" #144 *l(a)ut-?* 'fall or drop, follow, go or cross, run'. Bender also notes similarities in Afroasiatic (Chadic *-l*; not a good coincidence, presumably accidental) and Niger-Congo (*yRa, lyu* ...):

	NC	Ijoid	*la*
Cf.		Duru	*laa, luu* 'aller' (#91 in Boyd 1978: 69)

Blench (1998: 29) has proposed a similar innovation in Benue-Congo: 67.# *lo* 'to go' (Yoruba, Emai, Nupe, Yamba). Though word-initial lateral approximants are quite rare in most Nilo-Saharan languages (in the sample investigated here) this NS etymology is not very convincing. Important outside links were not observed. Nevertheless, it is a possible nuclear NS root. Cf. another 'to go'-root #37 in Gregersen (1972: 83) and *#ko* in Blench (1995: 116).

[5] Rare in Chadic, e.g. *as* in Sura (Jungraithmayr & Ibriszimow 1994, II: 106).

e#95 'grass' **LVØ** **(NS#68, CN#46, ES#57)**

Only 1 word-initial LVØ-occurrence was expected due to chance, but 3 cases were observed in the sample:

C	Maba	*lua*
E5	Nyimang	*lawa*
F	Mangbetu	*luɛ*

If transitional forms were included, the total is 5 occurrences in the sample languages; additionally:

B	Daza	*elle*	'grass, green'
E4	Gaam	*lel*	

Outside the sample:

E9	Shilluk	*ɛlɛi*	'a grass from which ropes are made'

Note also:

	Proto-Western Nilotic	**luum* 'kind of grass' (Rottland 1997: 154)
	Karimojong	*a-lɪɓa-nɪ* (Fleming 1983: 445) and
K	Ik	*elɪɓa* (contact with Nilotic)
E2	Tirma	*lanjɔi7* (Bender 1971: 265)
E8	Proto-Daju	**lasiəɣ* 'green', Thelwall's (1981: 178) reconstruction

This gloss was unused in Bender (1996a: 222, 160), though he mentions similar roots for the fragment 'grass, green'. Though this may be a common root for the nucleus, one possible explanation is mere sound symbolism (cf. 'lawn'). There is also a coincidence in East African Khoisan:

KS	Sandawe	*ɬ̰ípà*	'grass' (Elderkin 1983: 508)

e#112 'to kill 2'/'to die 2' ØVØ **(NS#81 'to kill 2', CN#30 'to die 2')**

This is another case where Greenberg's original etymology had to be split for semantic and statistical reasons - misleadingly (but necessarily) as is easy to see below. Four strict word-initial ØVØ-coincidences of 'to kill' were observed in the sample; 3 can be expected due to chance (closer to 4).

A	Gao	*wi*	

B	Daza	*i;*	*yidər, cido, yitər* 'tuer' (Le Cœur & Le Cœur 1956: 387)
C	Maba	*wu;*	*-ìyw-ír-* 'kill', *-ɔ́y-* 'die' in Edgar (1991b: 126, 128)
H	Kunama	*ya:*	

Additionally, 3 cases of 'to die' were observed in the sample, and also 3 word-initial ØVØ-coincidences were expected:

D	Fur	*wai*	
E4	Gaam	*iy;*	*tɪr* in Bender (1998: 56)
I	Koma	*ui*	(Madan); Komo *wíi, wu* 'die'; *k'ɔs (~š)* 'kill (Bender 1983: 269, 271)
			Note Shabo *ha* 'kill', *k'o* 'die' (Ehret 1995a: 186)

Outside the sample:

E9	Maasai	*ye;*	Proto-Nilotic **ɻɔ* 'die' (Dimmendaal 1988: 38)

Note the following supplementary evidence:

E5	Nyimang	*gwɛ́s-*	pl. 'to strike, kill', sg. *ní* (Tucker & Bryan 1966: 247)
E7	Tama	*iy*	'die', Ibiri=Merarit *yí/ey* 'kill' (Edgar 1991a: 126, 128)
L	Krongo	*aayá*	'die' (Schadeberg 1994: 26)

Ehret's (1989: 41, 43) roots: 'Sahelian' **wi* 'to kill' and 'Saharo-Sahelian' **yɛ* 'to die'. Cf. Bender's (1996a: 156, 185) most widespread Nilo-Saharan 'fragment' (possible "weak isogloss in Satellite-Core"), additionally in:

F	Central Sudanic *kui, wui*, Fc (Core Central Sudanic) **hwi*	

Besides, note also some similarity in Gumuz, e.g.:

J	Sai	*šəkwa*	'kill' (Bender 1979: 61)

Gregersen's (1972: 82, 84) Kongo-Saharan roots:
#24 'to die' (and Greenberg's NC root #14 'to die'), e.g.:

F	Lugbara	*gbī*	'kill', Lendu *hwi* 'kill';
NC	Huela	*kpã*	'kill'
	Bariba	*gbi*	

Kutep	*wẽ*	
Zande, Likpe	*kpi*	

and #44 'to kill':

NC	Kyan	*wuru*	'to die'
	Serer	*war*	'to die'
	Ewe	*wu*	
	Nupe	*wa*	
E1	Nubian	*iwire*;	Meidob *pɛrran*, Kadaru *wuri*, etc. (Thelwall 1978: 278)
E9	Turkana	*ari*	
F	Kreish	*iri*	
Cf. E2	Mursi	*ɛ́rrá*	'die' (Bender 1971: 265)
E8	Sila	*-irsi*	'die', Shatt *-ɐxsɵ* (Thelwall 1981: 176)

Westermann's (1927: 219, 225, 237-238) PWS included, with a wide occurrence of modern reflexes, the following reconstructions: **gue* 'töten'; **gù-* 'töten, sterben' (with reflexes *wu, wi*, etc.); **kú, *kúá, *kúí* 'sterben, töten'. See also Boyd's (1978: 56, 73) Adamawa-Ubangian roots for 'tuer' (#3 *i, yi*, etc.) and 'mourir' (#118 *wuu, wu*, etc.). Scattered similarities can be observed in:

AA	Qimant	*kʷɔ:*;	[Ehret's] Proto-Cushitic *kaɦ-* 'kill' (Bender 1994: 1162)
	Janjero	*wɔ:rɔ'wa* 'kill' (Bender 1971: 239, 258); Cf. Proto-Omotic **wUt'-* 'kill' (Bender 1994: 1156)	

Ehret (1995b: 478) has reconstructed a Proto-Afroasiatic root #1010 **yaʃ-* 'to die' (in practice only Cushitic?). Random coincidence may be seen in Khoisan:

KS	Sandawe	*kʔʷé:, wákʔʷà, húkʔʷà* 'to kill' (Elderkin 1983: 510, 515-516); Note Hadza *k||o-* 'to kill' (Sands 1995: 249)

This all seems very confusing. Because similar forms are found in both Nilo-Saharan and Niger-Congo, these two roots may at first sight be regarded as possible evidence for Kongo-Saharan. A possible explanation for the roots like *ya, yɛ* in NS might be borrowing from Cushitic (or a substrate). Concerning roots like *kui, hwi*, one possible source - in some cases at least - could be NC. Otherwise, the coincidences seem quite fragmentary. Still, many unanswered questions remain. Why are there such curious similarities in unrelated languages?

e#124 'man 1' ØVT (CN#61, ES#72)

Three word-initial ØVT-coincidences were expected and 4 were found in the sample languages, always strict. All these, except Berta, were Eastern Sudanic:

E1	Kenuz	*id*	
E2	Didinga	*ɛt*	
E3	Nera	*eite*	
G	Berta	*ide*;	*n'dɛ-* 'person' in Bender (1996a: 81)

Outside the sample, similar forms occurred in Central Sudanic:

F	Baka	*oda*,	Kreish *uddu*

Additionally, as supporting evidence we can find:

D	Lowland For	*dúɔ̀*	(Dornboos, unpubl./Bender)
E5	Nyimang	*waɋaŋ*	'person' (Stevenson 1957: 175)
E6	Temein(Ronge)	*ḍéénì*	'person', Doni *áḍɛ̀ŋ* (Stevenson, unpubl.)
E7	Sungor	*at*	'person', Ibiri *ìrtí* (Edgar 1991a: 129)

Cf. Bender's (1996a: 81, 105-106) 'Good isogloss' #22 **di* 'child or baby, brother, person, man, mother', "Mixed semantics: intertwined with other roots": 'Fair isogloss' #168 **t_2Ed-* (for t_2 see Bender 1996a: 69) 'boy or son or child, friend, man' and 'Satellite-Core' isogloss #176 **(a)ta* 'person, boy, child, father, man, son, sister'; 'person' additionally in the following:

B	Zagawa	pl. *otte*
I	Komo	*at(a)*
L	Krongo	*taa-/ɗɛ́ɛ́gɔ* 'man' (Talasa; L forms from Schadeberg 1994: 34)
	Mudo	*m-ɪɪɖɛ/kʸ-*
	Krongo	*kááú/kaḍú* 'person, people' (Krongo proper);
Cf. H	Ilit	*kaad-a*, Kunama *ka*, Bender (1971: 277-278)

Note also Afroasiatic (Omotic **AT* according to Bender 1996a: 106):

AA	Ari	*ɛ:d*, Male *a:si* (Bender 1971: 255, 263)
	Tigre	*ʔadda:m* (Bender 1971: 230)

Cf. Orel & Stolbova's (1995: 137-138) PHS #595 **da?-/*daw-* 'man, chief', with reflexes in Berber (**dVw-* 'men', e.g. *i-du* in Figig), Chadic (Musgum *dai* 'people')[6], and Rift (Cushitic; **daH-* 'stranger'); and Ehret's (1995b: 343) AA reconstruction #668: **ʃid-* 'person'; Omotic **id-*; Southern Omotic **ed-*. Accidental similarity in Khoisan and Niger-Congo:

KS	!Xóõ	*taa*	'person' (Sands 1995: 255)
NC	Proto-Bantu	**-jada*	(Meeussen 1980: 45)

Concerning this root, slight evidence for (nuclear) Nilo-Saharan might be observed. However, the distribution suggests that possibly an ancient contact with Afroasiatic languages can (speculatively) be seen as a possible source for it. Practically no connections with NC were found.[7] Further research is needed before any solid conclusions can be made.

e#128 'meat 2' NVØ/ØVN/NVN (NS#94, CN#64)

With strict criteria, there were 2 expected and 3 observed cases of word-initial NVØ-similarity in the sample:

C	Maba	*niu*;	*ñù-k* in Edgar (1991b: 128)
E3	Nera	*no*	
H	Kunama	*nya*	

In addition, the sample languages featured 5 transitional forms: 3 observed and expected ØVN- and 2 observed and expected NVN-sequences:

ØVN: B	Daza	*yini*
E4	Gaam	*ony*
G	Berta	*o:ŋ* (Fazoglo);
NVN: D	Fur	*neno, nino*
F	Mangbetu	*nyinyi*

The total is a remarkable 8 occurrences in the sample languages. Again, (possible) sound changes do not obey statistical restrictions. In addition:

[6] Note, however, the quite different roots in Jungraithmayr & Ibriszimow (1994, II: 230-231, 266-267).

[7] Cf. Greenberg's other different Nilo-Saharan 'man'-etymologies (*aba, bi*, etc. CN#62, NS#92), and Gregersen's (1972: 85) Kongo-Saharan roots #49 (*gur, kili, kalle*, etc.) and #50 (*boro, belu*, etc.).

E5	Nyimang	*nyiŋan*	(Ehret 1983: 411)
E6	Dese	*nányà?*	'oil' (Collective; Stevenson, unpubl.)
E7	Tama	*ŋan*	'eat' (Tucker & Bryan 1966: 208)
E9	Proto-Nilotic	**ɲam*	'eat, chew' (Dimmendaal 1988: 39)

Bender's (1996a: 90, 178) 'Good isogloss' #75 **Si(N)-* 'fat or oil, meat' overlapping with Ehret's (1989: 44) 'Saharo-Sahelian' root **yen* 'meat', the latter additionally in:

I	Komo	*yɛn*	'fat' (*šùm* 'meat', Bender 1983: 271)
K	Ik	*in*	'animal'

Compare with Bender's (1996a: 123) 'Symbolic' item #273, 'cat=leopard, lion, hyena, animal =meat=fat, fox' **-ɲa(u)-*; 'animal=meat=fat in:

H	Kunama	*(h)aɲa*
L	Krongo	*ɲááma* (Talasa 'animal, thing'; Krongo 'meat' *ǒǒɗa*; Schadeberg 1994: 22, 34, 43)

Note also Bender's (1996a: 124, 88) 'Good isogloss' #64 **Na* 'eat, bite, food, drink' and his 'Symbolic item' #275 **Nam* 'food, dura, eat or bite or burn'. Gregersen's (1972: 85) Kongo-Saharan root #51 for 'meat/animal'; e.g.:

NC	Mossi	*nem(do)*
	Ijo	*nama*
	Longuda	*nyɔmɔ*
	Proto-Bantu	**(n)yàmà*; **-(ɲ)ama* (Meeussen 1980: 45)
	Fulani	*nyaam-* 'to eat', *ko-nyaame* 'food'
A	Songai	*nyã* 'to eat'

Westermann's (1927: 269) PWS **-nìàm-, *-nàm-* 'Tier, Fleisch' with reflexes in Kwa, Kru, Benue-Congo, Ijoid and Gur. See also Boyd's (1978: 62) Adamawa-Ubangian roots for 'viande' (#43) *nài, ɲàmà, ɲā*, etc. Also found in some Chadic languages (rare):

AA	Hausa	*nāma* 'meat, flesh, (wild) animal'; etc. (Skinner 1996: 207)

Cf. also East African Khoisan:

KS	Sandawe	*|nin* (Kagaya 1993: 29)

This is a well-known 'common African root' with wide occurrence in both Nilo-Saharan and Niger-Congo. Because other African languages use different roots (with a handful

of exceptions), this might be a good piece of evidence for Kongo-Saharan. However, noteworthy is the often different place of articulation concerning the second nasal in Nilo-Saharan and Niger-Congo. This is another case where statistical methods lose transitional correspondences.

e#134 'mouth 2' **ØVK** **(ES#78 only)**

This etymology had, alongside the 4 observed strict word-initial ØVK-coincidences (3 expected), 3 quite different transitional forms ('doubtful judgments' according to Bender 1996b: 9). The total was 7 occurrences in 18 sample languages. All, except Nyangi, were Eastern Sudanic:

ØVK: E1	Kenuz	*agil*	
E4	Gaam	*ag*;	*ɔɬg, uɬg* in Bender (1996c: 145)
E8	Daju	*akkei*	
K	Nyangi	*ak*;	Heine's (1975: 295) Proto-Kuliak **ak* 'Mund'

Transitional forms:

E3	Nera	*aulo*;	Cf. *hagge* 'tongue' in Bender (1971: 268)
E5	Nyimang	*ŋal*;	*al* (Bender 1996c: 145)
E7	Merarit	*kul*;	Abuu Shaarib *awl* according to Edgar (1991a: 128)

This is one of Bender's (1996c: 145, 1998: 59) Ek/Eastern Sudanic -isoglosses (#2k 'mouth'): **(a)ŋgul*. Therefore, the etymology is not evidence for Nilo-Saharan, but instead for Eastern Sudanic, as it was in Greenberg (1966).[8]

---

[8] For another Nilo-Saharan root in Greenberg (1966), namely NS#96, CN#68 'mouth', strict coincidences were found only in 2 sample languages. With supplementary evidence there are many more, at least:

D	Fur	*udo*	
E2	Murle	*otok*	(Dimmendaal 1988:49)
E4	Sillok	*utu*	
E6	Temein	*ʊ̱Uk/kUʊ̱iŋ*	(Tucker & Bryan 1966: 256)
E9	Proto-Nilotic	**(k)UʊUk*	(Dimmendaal 1988:49)
F	Mongbutu	*uti*	
G	Berta	*(i)ndu*	(Bender 1998: 59)
H	Kunama	*u:da*	
I	Koma (Madin)	*tʔa*	(not in Bender 1996a: 135)

Cf. Bender's (1996a: 135) root #328 **(n)dO* 'mouth, tongue or language', **(k)utu(k)* in En; a partially similar form is also found in *NC *do* (*Mande *da* 'tongue'). Cf. Westermann's (1927: 246) **la* 'Mund' (reflexes in Kwa and Mande). See also Ehret (1981: 278 and 1983: 412-413).

e#144 'one' **TVK** **(NS#103, CN#72, ES#83)**

Strict cases of word-initial TVK-coincidence were found in 3 sample languages, 2 were expected:

C	Maba	*tek* (*tɔ́ɔ* in Edgar 1991b: 128; *duk* in Bender 1996a: 110)
D	Fur	*tok, dik*
E3	Nera	*doko, toko*

Additionally, outside the sample:

E9	Dinka	*tok*

In addition, one inexact similarity in the sample:

K	Nyangi	*odok*;	*nardok* in Heine (1975: 284); For a remarkable similarity see E8 and L below.

One semantically different form in the sample:

E7	Merarit	*tok*	'ten'

Similarly, outside the sample there is in Central Sudanic:

F	Dendje	*doko*	'ten'; Lendu *ɗi* 'one' in Fleming (1983: 457)

Note, as (at least weakly) supporting evidence:

E2	Mursi	*d'ò:nè*,	Murle *adoi* (Bender 1971: 265, 280)
E4	Gaam	*ḍɛwɛ-n*,	rare according to Fleming (1983: 457)
E5	Dinik	*ándà*	(Stevenson, unpubl.)
E6	Temein	*kíɗɔŋ*	(Stevenson 1991: 366)
E8	Liguri	*nɔhɔrɔk*	(Thelwall 1978: 279)
G	Berta	*duk'Una*	(Bender 1971: 269)
L	Talla	*ŋŋattɔk*,	Tulishi *kɔ́ttɔk*, Krongo *íŋwa* (Schadeberg =Kadugli 1994: 47)

Compare with Bender's (1996a: 110, 120) 'Satellite-Core isogloss' #199 **tUk* 'one, ten' and his 'Symbolic' item #261 **de(g)* 'one or first or alone or only, other, hand, two, four, ten'; 'one' additionally in:

I	Komo	*ɗe*

Partially similar is Westermann's (1927: 249) PWS **lé (dé, dó)* 'eins' with reflexes *do, de* etc. in Kwa, Kru and Mande. Gregersen's (1972: 85) Kongo-Saharan root #55:

NC	Malinke, Dan	*do*	
	Gwa	*dogbo*	
	Fõ	*ɖÒkpá*	
	Tamprusi	*dike*	
Note also:	Duru	*dáká*	'un' (#84 in Boyd 1978: 68)
	Bada	*ɗík*	'un' (Piron 1997: 500-503, possibly of Chadic origin)

Similar roots are found in some Afroasiatic languages (Chadic, Cushitic):

AA	Tangale	*dọk*	(Jungraithmayr & Ibriszimow 1994, II: 262)
	Oromọ	*tokko*	(Bender 1994: 1157)

See also Bender (1992: 24-26). This is a problematic root having a wide, and possibly non-accidental, occurrence in different African phyla. However, its distribution is quite unsystematic. Tentatively a few areal scatters might be outlined: one in West Africa, others in Ethiopia and central Africa. Similar roots are also found outside Africa. Note East African Khoisan:

KS	Sandawe	*ts'exe*	(Kagaya 1993: 48)

e#207 'tongue 1' **(KV)LVT** **(NS#140, CN#98, ES#116)**

Greenberg's lists included 2 sample languages (1 expected) with word-initial LVT-similarity:

E7	Merarit	*la:t*	
I	Koma	*litta*	(Kusgilo); Komo *lɛt'^{h}* in Bender (1983: 274)

In addition, there were 2 observed and 4 expected cases of word-initial KVL-sequence:

E4	Gaam	*kalat*	
G	Berta	*halad*	(Fazoglo); *hala~kala, xalaa* (Bender 1998: 62; En *kEl~kEɗ*)

Both these languages are spoken in the border area of Ethiopia and Sudan. (One of Bender's 1996a: 149 Nilo-Saharan fragments.) Additionally, Greenberg's etymologies included 3 other inexact forms in the sample languages:

D	Fur	*(d)ali*	
H	Kunama	*ŋe:la:*;	*kala* in Bender (1996a: 149)
F	Mangbetu	*kadra*	

Similarly, outside the sample:

E1	Garko	*jalde*; Meidob *kadaŋɪ*, Debri *ʃaldo*, in Thelwall (1978: 280)
E8	Dagu of Western Kordofan	*kuldaŋ*; Nyala *ʃabre* (Thelwall 1978: 281)
F	Lendu	*leda*

Possible supplementary evidence:

E5	Dinik	*ɔlɔ*	Note *ìlè~èlè* 'tooth' in Nyimang (Stevenson, unpubl.)
J	Hamej	*lɪta*	Other Gumuz dialects have somewhat different forms, e.g. Kokit *kwa:tɛta* (Bender 1979: 67)

However, cf. Greenberg's (1966: 63) Afroasiatic root #72, which is quite similar, e.g.:

AA	Hausa	*harše, halše*
	Berber	*ils*
	Sahidic Coptic	*las*, Ancient Egyptian *ns* (note *n~l*)
	Arabic	*lisa:n*

Orel & Stolbova's (1995: 361) reconstruction for PHS: #1666 **les* 'tongue', e.g. in Omotic **mi-las-*[9], and Semitic **lišān-*. Note also contact with Jarawan, e.g.:

NC	Bada	*lìs*	'langue' (Piron 1997: 254)

Several occurrences of (KV)LVT-coincidence in Nilo-Saharan were observed in and around Ethiopia. However, an important difference between NS and Afroasiatic can be observed: coronal stops in NS corresponding to the sibilants in AA. However, further research (involving semantic shifts like 'tongue'~'tooth/teeth'~'mouth') is necessary.

9 **t'iiL* in Bender (1994: 1157)

Possible explanations include contact and sound symbolism.[10] Compare with Bender's (1996a: 141) item #357 **TelN-* 'tongue or language, lick, taste, molar or teeth' - including F *(la)dra/lɛdrɛ, tɛra/taLa*, J *t'ɛt'a*, and I *lɛt'* etc. - "linking Nilo-Saharan and Afrasian" (Chadic **dlm* and Omotic **ts'il*).

e#209 'tooth' **NVK** **(ES#117 only)**

Only 2 word-initial NVK-similarities were expected due to random coincidence. Strict cases were found in 4 sample languages (all Eastern Sudanic, except Nyangi):

E1	Kenuz	*nihta, ni:ta*; Kenuz *nel*, Nobiin *niid*, etc. in Thelwall (1978: 280)
E2	Didinga	*nigitat/nigit*
E3	Nera	*nihi/nihitta*
K	Nyangi	*njik*, but see below

A transitional fifth form occurred in:

E4	Gaam	*niet* Because only the first alternative form in any sample language was taken into consideration, the second form (*nyigitu*) of Gaam in Greenberg's data was omitted from statistical consideration. However, according to Bender (1998: 62) it is *ɲiid*.

Note the following supplementary evidence (some weaker than others):

A	Dendi	*hínydyè* 'dent' (Zima 1994: 227)
C	Masalit	*káciŋgi*, Maba *sati-k* (Edgar 1991b: 131)
	Mimi (GD)	*ñaïn* 'dents' (Gaudefroy-Demombynes 1906: 154)
E5	Dinik	*ŋgíl/ŋgílià*; cf. Nyimang *ŋùl̪l̪ì* 'tongue' (Stevenson, unpubl.)
E6	Dese	*tyɬ̪k* (Stevenson, unpubl.)
E7	Tama	*ŋyìt* 'tooth, claw', Erenga *ɲɪlsit*, Ibiri *ŋóŋìt* (Edgar 1991a: 130)

[10] There is also another etymology for 'tongue (2)' in Greenberg (1966): NS#141:

B	Kanembu	*dələm*; Note Daza *teleši* 'langue' (Le Cœur & Le Cœur 1956: 331)
C	Maba	*delmi(k)*

Cf. Greenberg's (1963/1966: 23) NC#45 'tongue', Gregersen's (1972: 88) Kongo-Saharan root #74 'tongue', Blench's (1995: 113) Niger-Saharan *#deNe* 'tongue', and Westermann's (1927: 251) PWS **-lima* 'Zunge'.

E8	Shatt	*ʃix-te/ʃix-ke* (Stevenson 1991: 351)
	Nyala	*nyerte/nyerke* (Jungraithmayr 1978: 151)
F	Mangbetu	*nɛ'ki* (Larochette 1958: 228); *nɛ* being a prefix, 'déterminateur', according to Larochette. Could this be a re-analysis?
L	Krongo	*ʈ̪ɪn-jɪnɪ*, Kufo *nd̪iŋiní/ŋiini* (Schadeberg 1994: 43)

Ehret's (1989: 40) 'Kir-Abbaian/Astaboran' (E2, E3, E4) root **nikʰ* 'tooth'. According to Heine (1975: 300):

K	Nyangi	*eɬegw*/sg. *eɬegwód* 'Zahn'

Ehret (1981: 282) claims that Proto-Western-Kuliak **iɬVgw* is a loan from Cushitic. Similar are also some Nilotic, Central Sudanic, Kordofanian and Cushitic roots, e.g.:

E9	Maiak	*legit/lɛ:k*,Shilluk *lɛjɔ/lɛk* (Hall & Hall 1996: 161), etc.
F	Lendu	*lɛku* (Fleming 1983: 471)
Cf. NC	Moro	*l-əŋat/iŋaʈ̪* (Stevenson 1957: 147)
AA	Bilen	*ʔəlkʷi, eruk* in Orel & Stolbova (1995: 27); *ɨn'kwi:*,
	Somali	*ʔilig*, Konso *ɨlgɨta*; Alaba *In'kut* (Bender 1971: 238-247)

Cf. Orel's & Stolbova's (1995: 27) reconstructed PHS #103 **ʔilik-* 'tooth'; with clear reflexes in Cushitic. According to them: "The word for tooth looks like a Cush[itic] deverbative innovation and is preserved here because of the HS status of the corresponding verb." ['bite, chew']

Note:

AA	Egyptian	*nḥḏ* 'tooth, fang' (Old Kingdom), Orel & Stolbova's (1995: 273) PHS#1235 **ḥanVĉ-* 'tooth', claiming metathesis.
	Lele	*kāsīngá*; see other Chadic roots (Zaar *shin*, Higi-Nkafa *ɬɪnɛ*, Musgoy *ndin*, etc.) and the discussion in Jungraithmayr & Ibriszimow (1994, I: 170, II: 330-331).

Compare with the quite similar forms in Greenberg's (1966: 23, 159) Niger-Congo (Adamawa-Eastern) and Niger-Kordofanian lists (both etymology #46 'tooth'); e.g.:

NC	Mankanya	*(i)nyiŋ* (Mandyak)
	Bambara	*nyin*
	Dagomba	*nyine*
	Yoruba	*enyi*
	Daka	*nyine*;

	Talodi	*(j)Inyi*;
Cf.	Proto-Bantu	**-jíno* (Meeussen 1980: 53)

Westermann's (1927: 267-268) PWS **ní-, *-nín-* 'Zahn'; reflexes with wide occurrence (in Kwa, Kru, Benue-Congo, Gur, Atlantic, and Mande). See also Blench's (1995: 124) 'Niger-Saharan' root *#nyi* 'tooth'.

NVK-roots for 'tooth' are common in Eastern Sudanic, possibly being reconstructible for proto-ES. This root is usually missing in other NS lineages. A multitude of more or less similar forms can be found outside Nilo-Saharan. Does this mean borrowing, symbolism or what? At least some coincidences between NS and Afroasiatic may reflect ancient contacts (lateral approximants instead of nasals). Note that the root type generally is NV(N) in Niger-Congo, including the proto-language. The evidence seems to be too exhaustive for being only due to pure chance.

e#223 'white 2' **PVP[VL]** **(NS#150, ES#125)**

One expected and 2 observed cases of word-initial PVP-coincidence in the sample languages:

C	Maba	*fafara(k)*	
E5	Nyimang	*fɛfɛr*	'very white' [and 'bright', Ideophone];
		ṯabar	'white' (Stevenson 1957: 175 and 1991: 364)

Outside the sample:

E8	Sila	*papara*

Cf. Bender's (1996a: 111) 'Core-Group isogloss' #211 **fEr~~pur-* 'white, red or blood, yellow', 'white', e.g. in:

I	Twampa	*-p'er*
L	Krongo	*afiir-*

Lexical innovation #21n 'white, yellow' in En/Eastern Sudanic (also Ek-occurrence found in Nyimang) according to Bender (1996c: 147), e.g.:

E2	Murle	*fɔɔr*	
E4	Gaam	*bɔɔr*	
E6	Doni	*-fʊr*;	*afUrUnyà* 'yellow' (Stevenson, unpubl.)
E9	Lotuxo	*-b(w)or-*	
Cf.	Proto-Western Nilotic	**bor* according to Reh (as cited in Rottland 1997: 172)	

Note Greenberg's (1966: 23-24, 160) Niger-Congo (Adamawa-Eastern #49) and Niger-Kordofanian (#50) word lists, e.g.:

NC	Dyola	*fur*
	Sya	*foro*
	Mossi	*pel(ɣa)*
	Tiv	*pupu*
	Mumuye	*puru*
	Ngbandi	*vulu*;
	Masakin	*ipu*

Cf. Gregersen's (1972: 88) Kongo-Saharan root #78 'white', e.g.:

B	Kanuri	*bul*
NC	Avikam	*furu*
	Beri	*fufulu*

Blench's (1995: 113) 'Proto-Niger-Saharan' *#bulu* 'white'. Westermann's (1927: 279) PWS **pù-* 'weiß', with reflexes in Kwa, Kru, Benue-Congo, Gur, Atlantic, and Mande.

Compare also with:

F	Lugbara	*fɔrɔ-rɔ́* 'yellow' (Tucker 1940: 375)
AA	Gisiga	*babaraŋ* (Jungraithmayr & Ibriszimow 1994, II: 345)

Cf. Jungraithmayr & Ibriszimow's (1994, I: 178) Chadic root *pr*; restraining from reconstruction and assuming NC origin. An unimportant coincidence can be found in Omotic:

AA	Janjero	*fɔro*	(Bender 1971: 258)

Some support might be observed for an ES or nuclear NS root. However, this etymology may even be evidence for Kongo-Saharan. Yet sound symbolism, however, can be a plausible explanation in some cases at least. Note similarity with 'white 3'.

e#224 'white 3' **PVT** **(NS#151)**

The last of the three 'white'-etymologies in the combined list of Greenberg's Nilo-Saharan: the sample languages featured 4 strict cases of word-initial PVT-coincidence, and only 2 were expected with shared meaning:

D	Fur	*pota*

E4	Gaam	*po:den*; Note contact to Berta; *pɔɔ(n)* in Bender (1998: 63)
G	Berta	*fudi*
K	Nyangi	*bet*; *bɛj* Nyangi, **bec'* Proto-Kuliak in Heine (1976: 299)

Additionally, Greenberg's lists included one inexact form in the sample:

I	Koman	*mpata* (Kusgilo); cf. Gule *bit*

Bender's (1996a: 97) 'Fair isogloss' #120 **fVt* 'clean, rub, wash, white'.
In addition, there are 'regrettably' also other coincidences outside Nilo-Saharan:

NC	Grebo	*pudu, plu* 'to be white' (Westermann 1927: 279)
AA	Welamo	*bo:ts*^{a} (Elderkin 1982: 78)
	Amharic	*bič'a* 'yellow' (Fleming 1983: 473: "probably borrowed from Cushitic")
	Hausa	*fat* (ideo.) 'pure (white)' (Skinner 1996: 65)
KS	Sandawe	*p*^{h}*ó:* (Elderkin 1983: 510)
	Hadza	*pet*ɬ*'a* (Elderkin 1982: 78)

This type of root is quite rare in NS, and it is usually found in some languages of Ethiopia or nearby areas.

e#225 'who' NVØ (NS#152, CN#104, ES#126)

Eight cases of word-initial NVØ-coincidence were found in the sample of 18 languages; only 2 were expected due to chance. Only the first consonant is similar, the second being a zero consonant (i.e. short root ending in a vowel).

B	Daza	*nya*
C	Maba	*nyia*
E1	Kenuz	*ni*
E3	Nera	*na(n)*
E5	Nyimang	*ŋa*
E7	Merarit	*na*
E9	Bari	*ŋa*
H	Kunama	*na* 'who, which'

In addition, there was a 9th token, which was left out of statistical consideration due to the strictness of the criteria:

E2	Didinga	*ŋani*; Note Shabo *nee* 'who' (Ehret 1995a: 188)

Supplementary evidence:

E4	Gaam	*ŋɘna*	(Bender 1998: 63)
E6	Temein	*ŋa*, pl. *ŋa-ni*	(Stevenson 1957: 175)
E8	Liguri	*keneen*	(Thelwall 1978: 281)
G	Berta	*nda?*	(Bender 1998: 63)
L	Krongo	*ǹdá*	(Schadeberg 1994: 45)

Compare with Bender's (1991: 12, 23-24) 'interrogative formative' *ŋ*, additionally in:

F	Central Sudanic	*(i)ŋgo, nga, ɔŋɔ*, etc.

According to Gregersen (1972: 77) "Throughout both N[iger]-K[ordofanian] and NS, the stems of interrogatives and the interrogative suffix that often accompanies, have nasals." His examples include 'who' in:

NC	Proto-Bantu	**-nanì* or *-náà* (**nį* Meeussen 1980: 55)
	Ngbandi	*na*
Cf.	Duru	*nón* 'qui?' (Boyd 1978: 62), etc.

See also the partially similar roots for 'who' in several Omotic and some Cushitic languages:

AA	Konso	*ayno*
	Shinasha	*ko:ne*, Banna *ayne*, etc. (Bender 1971: 246-264)

This root was not used in Bender 1996a. It is well known that nasals are also used in interrogatives outside Africa. However, because it has such a wide occurrence in Nilo-Saharan, there might be some substance behind it.

5. Summary

As a summary, the following tentative sketch regarding Nilo-Saharan linguistic relationships can be proposed (see also Appendix 1). Quite robust evidence was observed for the nuclear Nilo-Saharan. The results suggest the revival and expansion of former 'Chari-Nile', corresponding approximately to Bender's 'Satellite-Core'. This nuclear Nilo-Saharan seems to include at least the following lineages:

- Maban, Foran, Eastern Sudanic, Central Sudanic, Berta, and Kunama

Also related to Nilo-Saharan in the light of this sample are:

- Kadu, and Koman

Whether these families belong to the nuclear Nilo-Saharan, cannot be decided from this sample. Besides the roots linking Kadu to Nilo-Saharan there are, however, some good coincidences between Kadu and Niger-Congo. Some clues were found demonstrating that the following lineages might have connections with Nilo-Saharan (possibly on a more remote level):

- Niger-Congo, and Saharan

Within this survey the lack of convincing evidence of their Nilo-Saharan affiliation appears to concern the following families:

- Songai, Gumuz, and Kuliak

Quite a few coincidences with Nilo-Saharan were found to occur in these stocks, but often similar roots were also observed outside Nilo-Saharan. Nevertheless, a more remote relationship is always possible, even though its demonstration might at least be difficult.

These results are based on a sample of lexical data, and the coincidences in grammatical morphemes can change the picture in some respects. More importantly, the evidence suggests that the effect of ancient loan words and possible substrate languages can remarkably distort the results of any attempt to classify distantly related languages.

6. Conclusion

Several clues, both quantitative (statistical) and qualitative (linguistic) for a remote relationship among almost all Nilo-Saharan lineages were observed. Unarguably, generally more evidence for Eastern Sudanic languages was found. The common origin for the nucleus of Nilo-Saharan seems plausible, possibly with Niger-Congo. However, the exact status of Nilo-Saharan remains unresolved until a comprehensive and systematic comparison with all Niger-Congo branches is conducted. Especially the position of Saharan languages remains obscure. The inclusion of questionable units, at least Songai and Kuliak, plus Gumuz, could not be substantiated.

Sometimes areal diffusion might be an explanation for the occurrence of similar roots in different phyla. Every root seems to have its own peculiar geographic distribution, quite frequently over the family borders. No single feature or root, whether cultural or not, can by itself prove linguistic relationship. Only the amount of systematic evidence beyond chance and its quality are of any relevance. Nevertheless, even spuriously significant coincidences can be found, these being of no comparative value. Presumably the role of sound symbolism has also been underestimated.

Acknowledgements

I would like to express my gratitude to M. Lionel Bender and Roger Blench for kindly making available unpublished material (especially the Temein and Nyimang word lists collected by Roland C. Stevenson), and Margot Whiting for language revision.

References

Bender, M. L. 1971.
The Languages of Ethiopia. **Anthropological Linguistics** 13(5): 165-288.

1979 *Gumuz: A Sketch of Grammar and Lexicon.* **Afrika und Übersee** LXII(1): 38-69.

1981 Some Nilo-Saharan Isoglosses. In: *Nilo-Saharan. Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980*, Thilo C. Schadeberg & M. Lionel Bender (eds.), pp. 253-267. Dordrecht: Foris.

1983 *Proto-Koman Phonology and Lexicon.* **Afrika und Übersee** LXVI(2): 259-297.

1991 Subclassification of Nilo-Saharan. In: *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30 - Sep. 2, 1989,* M. Lionel Bender (ed.) (=Nilo-Saharan Linguistic Analyses and Documentation 7.), pp. 1-35. Hamburg: Buske.

1992 *Classification génétique des langues Nilo-Sahariennes.* **Linguistique Africaine** 9: 15-39.

1994 The Mystery Languages of Ethiopia. In: *New Trends in Ethiopian Studies. Papers of the 12th International Conference of Ethiopian Studies. Michigan State University 5-10 September 1994. Volume I: Humanities and Human Resources*, Harold G. Marcus & Grover Hudson (eds.), pp. 1153-1174. Suite D Lawrenceville: The Red Sea Press.

1996a *The Nilo-Saharan Languages: A Comparative Essay.* München: Lincom Europa.

1996b Nilo-Saharan '95. In: *Proceedings of the Sixth International Nilo-Saharan Linguistics Conference, Santa Monica 1995, March 27-29*, M. Lionel Bender & Thomas J. Hinnebusch (eds.) (=**Afrikanistische Arbeitspapiere** 45.), pp. 1-25.

1996c Genetic Sub-grouping of East Sudanic. In: *Proceedings of the Sixth International Nilo-Saharan Linguistics Conference, Santa Monica 1995, March 27-29*, M. Lionel Bender & Thomas J. Hinnebusch (eds.) (=**Afrikanistische Arbeitspapiere** 45.), pp. 139-150.

1998 *The Eastern Jebel Languages of Sudan II. Comparative Lexicon.* **Afrika und Übersee** 81(1): 39-64.

Bender, M. L. (ed.) 1989.

Topics in Nilo-Saharan Linguistics. (=Nilo-Saharan Linguistic Analyses and Documentation 3.) Hamburg: Buske.

1991 *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30 - Sep. 2, 1989*. (=Nilo-Saharan Linguistic Analyses and Documentation 7.). Hamburg: Buske.

Bender, M. L. & T. J. Hinnebusch (eds.) 1996.
Proceedings of the Sixth International Nilo-Saharan Linguistics Conference, Santa Monica 1995, March 27-29. (=**Afrikanistische Arbeitspapiere** 45.)

Bendor-Samuel, J. (ed.) 1989.
The Niger-Congo Languages: A classification and description of Africa's largest language family. Lanham: University Press of America.

Blench, R. 1995.
Is Niger-Congo Simply a Branch of Nilo-Saharan? In: *Cinquième Colloque de Linguistique Nilo-Saharienne/Fifth Nilo-Saharan Linguistic Colloquium, Nice, 24-29 août 1992, Actes /Proceedings,* Robert Nicolaï & Franz Rottland (eds.) (=Nilo-Saharan 10), pp. 83-130. Köln: Köppe.

1998 *The Benue-Congo Languages: A Proposed Internal Classification.* Working document, Cambridge.

Boyd, R. 1978.
A propos des ressemblances lexicales entre langues Niger-Congo et Nilo-Sahariennes. In: *Études Comparatives: Oubanguien et Niger-Congo - Nilo-Saharien.* (=**Bibliothéque de la SELAF** 65.), Raymond Boyd & France Cloarec-Heiss (eds.), pp. 43-94. Paris.

1996 Congo-Saharan revisited. In: *Afrikanische Sprachen zwischen Gestern und Morgen: Beiträge zur Dokumentation, Klassifikation und Rekonstruktion,* Uwe Seibert (Hrsg.) **Frankfurter Afrikanistische Blätter** 8: 15-48. Köln: Köppe.

Boyd, R. & F. Cloarec-Heiss (eds.) 1978.
Études Comparatives: Oubanguien et Niger-Congo - Nilo-Saharien. (=**Bibliothéque de la SELAF** 65.). Paris: Société d'Études Linguistiques et Anthropologiques de France.

Cyffer, N. 1994.
English - Kanuri Dictionary. (Westafrikanische Studien 3.) Köln: Köppe.

Dimmendaal, G. J. 1988.
The Lexical Reconstruction of Proto-Nilotic: A First Reconnaissance. **Afrikanistische Arbeitspapiere** 16: 5-67.

Edgar, J. 1991a.
First steps toward Proto-Tama. In: *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30 - Sep. 2, 1989*, M. Lionel Bender (ed.) (=Nilo-Saharan Linguistic Analyses and Documentation 7.), pp. 111-131. Hamburg: Buske.

1991b *First Steps Toward Proto-Maba.* **African Languages and Cultures** 4(2): 113-133.

Ehret, C. 1981.
The Classification of Kuliak. In: *Nilo-Saharan. Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980,* Thilo C. Schadeberg & M. Lionel Bender (eds.), pp. 269-289. Dordrecht: Foris.

1983 Nilotic and the Limits of Eastern Sudanic: Classificatory and Historical Conclusions. In: *Nilotic Studies.* Part Two, Rainer Vossen & Marianne Bechhaus-Gerst (eds.) (=Kölner Beiträge zur Afrikanistik 10.), pp. 375-421. Berlin: Reimer.

1989 Subclassification of Nilo-Saharan: A Proposal. In: *Topics in Nilo-Saharan Linguistics*, M. Lionel Bender (ed.) (=Nilo-Saharan Linguistic Analyses and Documentation 3.), pp. 35-49. Hamburg: Buske.

1995a Do Krongo and Shabo Belong in Nilo-Saharan? In: *Cinquième Colloque de Linguistique Nilo-Saharienne/Fifth Nilo-Saharan Linguistic Colloquium, Nice, 24-29 août 1992, Actes /Proceedings*, Robert Nicolaï & Franz Rottland (eds.) (=Nilo-Saharan 10), pp. 169-193. Köln: Köppe.

1995b *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. **University of California Publications in Linguistics** 126. Berkeley: University of California Press.

Elderkin, E. D. 1982.
On the Classification of Hadza. **Sprache und Geschichte in Afrika** 4: 67-82.

1983 Tanzanian and Ugandan Isolates. In: *Nilotic Studies,* Part Two, Rainer Vossen & Marianne Bechhaus-Gerst (eds.) (=Kölner Beiträge zur Afrikanistik 10.), pp. 499-521. Berlin: Reimer.

Fleming, H. C. 1983.
Kuliak External Relations: Step One. In: *Nilotic Studies,* Part Two, Rainer Vossen & Marianne Bechhaus-Gerst (eds.) (=Kölner Beiträge zur Afrikanistik 10.), pp. 423-478. Berlin: Reimer.

Gaudefroy-Demombynes, M. 1906.
Documents sur les langues de l'Oubangui-Chari. (Extrait du tome II des Actes du XIV^{e} Congrès International des Orientalistes). Paris: Ernest Leroux.

Greenberg, J. H. 1966.
The Languages of Africa. Bloomington: Indiana University. Third Edition. (Originally published as Part II of the **International Journal of American Linguistics** 29, 1, 1963.)

Gregersen, E. A. 1972.
Kongo-Saharan. **Journal of African Languages** 11(1): 69-89.

Hall, B. L. & R. M. R. Hall. 1996.
From Cognate Set to Etymology in the Reconstruction of Proto-Nilotic. In: *Proceedings of the Sixth International Nilo-Saharan Linguistics Conference, Santa Monica 1995, March 27-29*, M. Lionel Bender & Thomas J. Hinnebusch (eds.) (=**Afrikanistische Arbeitspapiere** 45.), pp. 151-175.

Heine, Bernd. 1975.

Tepes und Nyang'i: zwei ostafrikanische Restsprachen. **Afrika und Übersee** LVIII(3/4): 263-300.

1976 *The Kuliak Languages of Eastern Uganda.* Nairobi: East African Publishing House.

Jungraithmayr, H. 1978.
A Lexical Comparison of Darfur and Wadai Daju. *Aspects of Language in the Sudan.* Robin Thelwall, (ed.) (= Occasional Papers in Linguistics and Language Learning 5.), pp. 145-154. Coleraine: The New University of Ulster.

Jungraithmayr, H. & D. Ibriszimow 1994.
Chadic Lexical Roots. 2 vols. (=Sprache und Oralität in Afrika 20.) Berlin: Reimer.

Kagaya, R. 1993.
A Classified Vocabulary of the Sandawe Language. (=Asian & African Lexicon 26) Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).

Larochette, J. 1958.
Grammaire des Dialectes Mangbetu et Medje. **Annales du Musée Royal du Congo Belge, Série in 8°, Science de l'Homme, Linguistique** 18. Tervuren.

Le Cœur, Ch. & M. 1956.
Grammaire et textes Teda-Daza. *Mémoires de l'Institut français d'Afrique noire* 46. Dakar: Ifan.

Marcus, H. G. & G. Hudson (eds.) 1994.
New Trends in Ethiopian Studies. Papers of the 12th International Conference of Ethiopian Studies. Michigan State University 5-10 September 1994. Volume I: Humanities and Human Resources. Suite D Lawrenceville: The Red Sea Press.

Meeussen, A.E. 1980.
Bantu lexical reconstructions. (Reprint) **Archives d'Anthropologie** 27. Tervuren: Musée Royal de l'Afrique Centrale.

Mikkola, P. 1998.
Random Coincidence in Mass Comparison: Preliminary Analysis of the Nilo-Saharan Lexicon. **Nordic Journal of African Studies** 7(1): 63-92.

Nicolaï, R. & F. Rottland (eds.) 1995.
Cinquième Colloque de Linguistique Nilo-Saharienne/Fifth Nilo-Saharan Linguistic Colloquium, Nice, 24-29 août 1992, Actes/Proceedings. (=Nilo-Saharan 10). Köln: Köppe.

Orel, V. E. & O. V. Stolbova. 1995.
Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction. (=Handbuch der Orientalistik/Handbook of Oriental Studies: Erste Abteilung. Der Nahe und Mittlere Osten/The Near and Middle East 18.) Leiden: E. J. Brill.

Piron, P. 1997.
Classification interne du groupe bantoïde. 2 vols. München: LINCOM Europa.

Rottland, F. 1997.
Lexical Reconstructions within Nilotic. **Afrikanistische Arbeitspapiere** 52: 139-173.

Sands, B. E. 1995.
Evaluating Claims of Distant Linguistic Relationships: The Case of Khoisan. PHD diss. Los Angeles: University of California.

Schadeberg, T. C. 1994.
Comparative Kadu Wordlists. **Afrikanistische Arbeitspapiere** 40: 11-48.

Schadeberg, T. C. & M. L. Bender (eds.) 1981.
Nilo-Saharan. Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980. Dordrecht: Foris.

Skinner, N. 1996.
Hausa Comparative Dictionary. (=Westafrikanische Studien 11.) Köln: Köppe.

Stevenson, R. C. 1957.
A Survey of the Phonetics and Grammatical Structure of the Nuba Mountain Languages. With Particular Reference to Otoro, Katcha and Nyimaŋ. **Afrika und Übersee** XLI, 3: 117-152; 4: 171-196.

1991 Relationship of Kadugli-Krongo to Nilo-Saharan: Morphology and Lexis. In: *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30 - Sep. 2, 1989*, M. Lionel Bender (ed.) (=Nilo-Saharan Linguistic Analyses and Documentation 7.), pp. 347-369. Hamburg: Buske.

Thelwall, R. 1978.
Lexicostatistical Relations between Nubian, Daju and Dinka. In: *Études Nubiennes: Colloque de Chantilly, 2-6 juillet 1975.* (=Bibliothèque D'Étude LXXVII, Institut Français d'Archaéologie Orientale du Caire.), pp. 265-286.

1981 Lexicostatistical Subgrouping and Lexical Reconstruction of the Daju Group. In: *Nilo-Saharan. Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980*, Thilo C. Schadeberg & M. Lionel Bender (eds.), pp. 167-184. Dordrecht: Foris.

Thelwall, R. (ed.) 1978.
Aspects of Language in the Sudan. (=Occasional Papers in Linguistics and Language Learning 5.) Coleraine: New University of Ulster.

Tucker, A. N. 1940.
The Eastern Sudanic Languages I. London: Oxford University Press.

Tucker, A. N. & M. A. Bryan 1966.
Linguistic Analyses: The Non-Bantu Languages of North-Eastern Africa. London: Oxford University Press.

Vossen, R. & M. Bechhaus-Gerst. (eds.) 1983.
Nilotic Studies. Proceedings of the International Symposium on Languages and History of the Nilotic Peoples, Cologne, January 4-6, 1982. (=Kölner Beiträge zur Afrikanistik 10.) Berlin: Reimer.

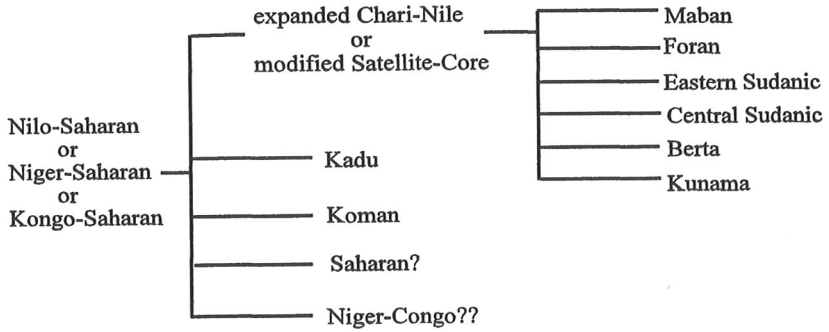
Westermann, D. 1927.

Die westlichen Sudansprachen und ihre Beziehungen zum Bantu. (=**Beiheft zu den Mitteilungen des Seminars für Orientalische Sprachen** XXIX.) Berlin: de Gruyter.

Williamson, K. 1989.
Niger-Congo Overview. In: *The Niger-Congo Languages: A classification and description of Africa's largest language family*, Bendor-Samuel, John (ed.), pp. 3-45. Lanham: University Press of America.

Zima, P. 1994.
Lexique Dendi (Songhay): (Djougou, Bénin). (=Westafrikanische Studien 4.) Köln: Köppe.

Apperndix 1. Tentative family tree (working hypothesis only!).



Possibly independent families: Kuliak
Songai
Gumuz

Appendix2. Observed coincidences in Nilo-Saharan stocks and some control lineages, including only those sample etymologies featuring more observed than expected tokens in Greenberg's word lists.

Stock and survey language	# of etymology in the combined list													
	60	91	95	112	124	128	134	144	207	209	223	224	225	Total[1]
A Songai	■			■		▽				○				1- 4
B Daza		▽	■	■	○	■					○		■	2- 7
C Maba	▽		■	■		■		■		○	■		■	6- 8
D Fur	■			■	●	■	▽	■	■			■		3- 8
E1 Kenuz				▽	■		■		□	■	▽		■	4- 7
E3 Nera	■	■			■	■	■	■		■	▽		■	6- 9
E5 Nyimang			■	●	●	●	■	▽	▽	○	■		■	3-10
E7 Merarit	■	■		●	○	▽	■	■	■	●			■	4-10
E2 Didinga			○	▽	■		▽	○		■	○		■	2- 8
E4 Gaam			■	■		■	■	●	■	■	●	■	●	2-10
E6 Temein					●	▽	▽	●		○	○		●	0- 7
E8 Daju	■		▽	▽			■	▽	□	●	□		▽	2- 9
E9 Bari	○	□	□	□		▽	▽	□		○	○		■	1-10
F Mangbetu	■	□	■	○	□	■	▽	□	■	●			▽	1-11
G Berta					■	■	▽	●	■			■	●	2- 7
H Kunama		■		■	○	■	▽		■		▽		■	4- 8
I Koma				■	●	●	▽	●	■		○	■		1- 8
J Gumuz				▽					●					- 2
K Nyangi	▽		▽			▽	■	■		■		■		3- 7
L Krongo	○			●	●	○		○		●	●		●	- 8

Coincidences in major control lineages (preliminary investigation)														
NC		∇		●		●		∇		●	●		∇	- 7
Omotic	∇			∇	●								∇	- 4
Cushitic	●							∇		●				- 3
Chadic								∇	●	∇	∇			- 4
Sandawe			●	∇		∇		●				●		- 5
Explanations	etymology in the combined list: number and gloss													
	60 dog	**91** go l	**95** grass	**112** kill 2	**124** man 1	**128** meat 2	**134** mouth 2	**144** one	**207** tongue 1	**209** tooth	**223** white 2	**224** white 3	**225** who	

1 The first figure of the total indicates the number of strict coincidences in the original sample, the second includes other observed similarities also (such as transitional coincidences, related languages, etc.).

Note on symbols:

■ strict word-initial CVC-coincidences in the original sample
■ transitional coincidences in the original sample (including semantically different forms, like e#112 'die 2')
□ Greenberg's evidence outside the sample

Supplementary evidence from other sources (not necessarily strict coincidences):

● sample languages (plus Gumuz and Krongo), and wide occurrence in the major control lineages
○ other languages within the same (Nilo-Saharan) stocks
∇ Other possible cognates or random coincidences (somewhat subjectively chosen, not exhaustive), including less common or more divergent forms; single and scattered occurrences in control languages are not included. If more semantic shifts are accepted, even more possible cognates might be found.

Nordic Journal of African Studies 8(2): 139-157 (1999)

SALAMA
SWAHILI LANGUAGE MANAGER

ARVI HURSKAINEN
University of Helsinki

SALAMA is an acronym for *Swahili Language Manager*[1]. It is a computerised working environment, where it is possible, with the help of a set of programs and user-defined utilities, to perform a multitude of tasks. For introducing SALAMA, it is perhaps more interesting to describe its aims and applications first, and then give an outlay of its components. Appendix 1 gives an overall view of the structure of the system, and of some of its applications. In this paper it is possible to give only a brief and condensed description of how the system is constructed. It is without any technical detail as to how the system is implemented. For those interested in more detailed description of the components, a list of relevant publications is added in the end of the paper. References to those are also made in text.

WHY SALAMA?

SALAMA has grown gradually from a rudimentary morphological parsing program into a comprehensive language management system. In other words, there was no original plan for creating SALAMA. The accomplishment of one phase has given impetus to another effort, and by building one block on another, an end product of formidable properties has resulted.

At present, SALAMA has facilities for carrying out such task as:
Spelling checker of Standard Swahili text (implemented on Word 97 and later versions)
Hyphenator for automatic hyphenisation of Swahili text (implemented on Word 97 and later versions)
Morphological analyser, with information on such features as: part-of-speech (word class), tags for inflectional and derivational morphemes, lemma, etymology of loan-words, tags for domain-specific terminology, gloss in English, etc.
Lemmatiser

1 SALAMA was introduced and demonstrated for the first time in the 20th International Biennial Conference of the African Language Association of Southern Africa (ALASA), in July 5-9, 1999.