Studies in African Linguistics
Volume 17, Number 1, April 1986

THE LEXICOSTATISTIC BASE OF BENNETT & STERK'S RECLASSIFICATION OF NIGER-CONGO WITH PARTICULAR REFERENCE TO THE COHESION OF BANTU*

Thilo C. Schadeberg
Rijksuniversiteit te Leiden

In 1977, Bennett and Sterk published a reclassification of the Niger-Congo languages which has been highly influential. In this paper I try to discover their lexicostatistic method (section 1), then use their published data to do a conventional lexicostatistic subgrouping (section 2), and finally look at their evidence for denying the genetic unity of Narrow Bantu (section 3).

1. Bennett and Sterk's Method

Bennett and Sterk's lexicostatistic method is not fully described in their 1977 paper: "A full account of the procedures followed and their theoretical justification is being prepared for publication elsewhere" (p. 242). Since this full account has to my knowledge not yet appeared, and since they obviously use new methods which they developed themselves, some interpretation is necessary.

Bennett and Sterk used a "computer-aided weighted count study" (p. 242). The weighting seems to have consisted of a three-level cognate scoring: Level 1 (the most "generous" one) counts every likely cognate; at Level 2 cognate sets may be split into several sets on the basis of variations (they provide the example lem vs. mel 'tongue'); at Level 3 even finer details (such as noun classes) are distinguished. In practice, however, only Level 1 provided

*I wish to thank Kay Williamson for her helpful comments on a draft of this small paper. She also was so kind as to let me use a file which she had been given by Jan Sterk, containing data and notes that were used for the Bennett and Sterk [1977] article. I am grateful for this chain of generous cooperation which helped me to a better understanding of how Bennett and Sterk reached their important reclassification of Niger-Congo.

useful results since already at Level 2 most relationships fell below their cut-off point of 18%. It therefore remains unclear how much "weighting" actually entered their lexicostatistics. (The similarity matrix corresponding to their Level 1 cognate scoring is reproduced in their article.)

Bennett and Sterk augmented their lexicostatistic study with a search for group specific innovations. "Where the two types of study disagreed, the innovation-based evidence was given preference" (p. 245). I shall briefly return to the proposed innovations in section 3 in as far as they concern Bantu.

Tree-generating lexicostatistics is based on hierarchical cluster anaysis. Bennett and Sterk use two devices which make straightforward hierarchical analysis impossible. The first one is their use of blanks for all scores of less than 18%. I think one is right to disregard values below 20%, just as I would not use this kind of lexicostatistics to classify a language group in which most members score more than 80% cognates. However, in order to calculate hierarchical clusters a blank as such is not a possible input. It has to be interpreted as some value, possibly even zero. In my own study I have decided to interpret Bennett and Sterk's blanks as representing the value 17%. Hence, my results say nothing about those most remote relationships, which is exactly what Bennett and Sterk and I want. Interpreting blanks as zero or some intermediate value would lead to gross and undesirable distortions in the calculations of branch averages.

The other feature which is unsuitable for hierarchical cluster analysis is that two figures are provided for each pair of languages. In other words, the distance between language A and language B is not necessarily the same as the distance between language B and language A. This is the result of Bennett and Sterk's way to handle blanks of which there are two kinds. The first kind simply represents missing entries. The other kind of blank arises when one language has two entries for one meaning and the other has only one. Suppose we have four words in two languages:

	A	B	
ear	1	1	
eat	1	1,2	
egg	1	2	
eye	0	1	[0 = no entry]

B shares two of the three words in language A (67%), but A only shares two of the five words in B (40%). If that is what Bennett and Sterk have done then languages with complete lists, i.e. few gaps, should consistently score lower than languages with less complete lists. Such languages do exist, e.g. Kikuyu and Tiv. Since there are quite a few cases where the distance A:B differs by ten or more points from the distance B:A I fear that for some languages the available lists contained rather more gaps than is desirable for any lexicostatistics.

Since I think one should base cognation percentages on the number of comparisons rather than words, I have decided to use for each pair of languages the higher of Bennett and Sterk's figures. The underlying assumption is that if the blank were filled in the item would have the same likelihood of being cognate as the average likelihood of all other items taken together. This may not be quite true if different words have different likelihoods of being replaced in the course of time (cf. Dyen, James and Cole [1967]) and if in addition short wordlists are more likely to contain more stable words than less stable ones. It is a purely subjective impression of my own that the last condition may be true. A wordlist containing the less stable item 'leaf' will almost certainly also contain the more stable item 'tree', whereas the inverse does not hold. Still, as long as the number of missing items is small the most common and quite acceptable method is to base the percentage of cognates solely on the number of actual comparisons.

2. A Pure Lexicostatistic Subclassification[1]

The two extreme methods for hierarchical subclassification are the Nearest

[1] The lexicostatistic calculations used for this paper were carried out with the program LEXISTAT. I have written this program in Pascal, to run on

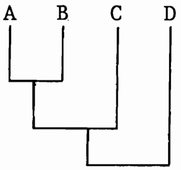
Neighbour (NN) and the Furthest Neighbour (FN) methods. They differ in what they take to be the distance (cognation percentage) between a cluster X and another cluster or language Y. NN assumes that the distance is equal to the closest distance between any member of X and (any member of) Y; FN takes the greatest distance as its measure. This can lead to competing clusterings when four or more languages are being classified. A hypothetical example will help to clarify the difference between NN and FN:

	A	B	C	D
A	-			
B	60	-		
C	50	40	-	
D	35	40	45	-

Nearest Neighbour

	AB	C	D
AB	-		
C	50	-	
D	40	45	-

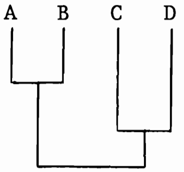
	ABC	D
ABC	-	
D	45	-



Furthest Neighbour

	AB	C	D
AB	-		
C	40	-	
D	35	45	-

	AB	CD
AB	-	
CD	35	-



IBM PC and compatible computers with PC-DOS or MS-DOS. LEXISTAT accepts either a table of cognation judgements or a similarity matrix as its input. It carries out several lexicostatistic analyses; it allows selective use of the cognation judgement table and the deletion of specified languages from particular cluster analyses. It produces tabular and graphic results. I would be happy to share this program with anyone who is willing to compensate me for the price of the diskette plus postage.

If the assumptions underlying lexicostatistics were fully correct, and if words were never borrowed between related languages (or could always be detected as such) then both methods should provide identical results. Unfortunately they seldom do. Nearest Neighbour (NN) typically produces "onion type" trees, i.e. a succession of splits between one or a few language(s) on one side as against the rest of the languages on the other side. Furthest Neighbour (FN) tends to produce more balanced trees. In principle, FN should be less distorted by borrowing between part of the languages of one branch and part of the languages of another branch. Various methods exist that mediate between NN and FN by taking various types of averages as the distance between clusters. That means that any node that appears in both extreme methods will also appear in any averaging method. Figures 1, 2, and 3 (in the Appendix) show the trees resulting from Branch Average (BA), NN, and FN subclassification. Table 2 gives the corresponding figures, and Table 3 contains the revised similarity matrix.

Accepting for the time being the reliability of the basic data I suggest interpreting these trees in the following way. First, let us accept all nodes that are common to both the NN and the FN trees. Then, on a somewhat lower level of confidence, let us accept the nodes that the BA tree shares with either the FN or the NN tree and that are not strongly contradicted by the "opposite" tree. The reasoning for this is that while FN, in principle, is most likely to produce genetic trees, both NN and FN are particularly sensitive to distortion by poor data, either primary or by wrong cognation judgements; this is where BA comes in as a corrective. In this way we may arrive at the following conclusions. There appear to be nine primary branches, and the largest of these may be divided into nine secondary branches (see list on following page). Branches marked with an asterisk represent nodes that are stable between NN and FN. Unmarked branches are less strongly supported. "(New) Kwa" represents Bennett and Sterk's "Western SCNC", i.e. the old Western Kwa. "(New) Benue-Congo" represents Bennett and Sterk's "Eastern SCNC", i.e. old Eastern Kwa plus Benue-Congo. According to the NN classification, (New) Kwa lacks internal unity presumably because a few figures have been inflated by

1. Fula*
2. Dyola*
3. Temne*
4. Kru*
5. Gur*
6. Adamawa-Ubangi (?)
7. (New) Kwa
8. Ijo*
9. (New) Benue-Congo

9.1 Nupoid*
9.2 Idomoid*
9.3 Yoruboid*
9.4 Edoid*
9.5 Igbo(id)*
9.6 Jukunoid*
9.7 Cross-River
9.8 Plateau (?)
9.9 Bantoid

areal contact. (New) Benue-Congo falls into three distinct branches in the FN classification; this is entirely due to a few scattered cognation scores below 18%. Adamawa-Ubangi has been marked as doubtful because it is only supported by the FN classification; in the BA classification, Tula clusters with the Gur languages and creates a link between Gur and Adamawa-Ubangi.

As far as the "primary" branches are concerned, our results do not disagree with those reached by Bennett and Sterk, though the 18% cut-off obliterates any possible evidence for the more detailed tree structure which they propose on different grounds.

The first six subbranches of (New) Benue-Congo are lexicostatistically stable between NN and FN subclassifications. The internal unity of Cross-River is not supported by NN because of the curiously low cognation scores between Efik and the other two representatives of this branch. Plateau is marked as doubtful, but in fact only the inclusion of Kambari is doubtful. Finally, Bantoid as a whole is not supported by NN because the non-Bantu Bantoid languages Tiv, Mambila, and Jarawan have individually varied affiliations within (New) Benue-Congo.

In summary then, lexicostatistics supports groupings rather similar to those proposed by Bennett and Sterk for their South-Central Niger-Congo, though the tree has less internal structure and notably lacks the intermediate nodes Central Niger and Benue-Zambesi.

3. The Internal Cohesion of Bantu

We have already found that Bantoid appears to be a lexicostatistically valid branch of (New) Benue-Congo since it appears in both the FN and the BA cluster analysis. In addition it must be observed that the internal structure of this branch is almost identical in both analyses, in particular the primary subdivision between non-Bantu Bantoid and (Narrow) Bantu. Moreover, (Narrow) Bantu is a stable node which appears not only in FN and BA but also in the NN tree. It would be unwise to base an internal subclassification of Bantu on the five languages represented in this study, but it must further be noted that there is no lexicostatistical evidence here to support the subdivision into "Equatorial" (Northwest Bantu: zones A, B, C, and part of D) and "Zambesi" (the remainder). Therefore, the present figures provide no support at all for the proposal by Bennett and Sterk that "the greatest departures from previous classifications lie ... among the Bantoid languages, now grouped under the heading Benue-Zambesi, where Guthrian Bantu does not appear to constitute a valid subgrouping" (p. 241).

I assume then, that the proposed disintegration (rather than just subclassification) of Bantu rests solely on (non-)shared innovations. Bennett and Sterk propose three isoglosses separating "Ungwa" (= Zambesi Bantu plus Tiv) from "Wok" (= Equatorial Bantu, Ekoid, and Mbam-Nkam plus Jarawan). Two of these isoglosses are defined as innovations: "Ungwa" has ungwa 'hear' where "Wok" has preserved wɔk , and "Wok" has -ɔŋ 'hair' where "Ungwa" has preserved SCNC nyúélé . The third isogloss concerns an item -baŋ 'red' which is found only in "Wok" (p. 261). The two innovations ('hear' and 'hair') may well refer to complex sound shifts, not to simple lexical isoglosses. The exact correspondences for these lexical items have not yet been worked out for (Narrow) Bantu.

Meeussen [1980] reconstructs *-jį́gų- 'hear' and notes uncertainty about the first vowel (į/i/u) , the second vowel (ų/u) , and the medial consonant (g/ŋg..) . Guthrie's Common Bantu also contains -yį́(n)g(ų)- and -yú(n)g(ų)- (plus some other variants). Bennett and Sterk's form wɔk is the equivalent of Guthrie's Bantu form -yúg- . The problem is complex because this verb is highly peculiar in its phonological make-up; it combines all the most difficult segment sequences in a rare, non-canonical shape. Since it is likely that all these forms are ultimately cognate, the real innovation could only be one of the sound shifts separating these forms. Zambesi Bantu attests both front and back vowels as V_1, and prenasalized as well as simple g as C_2. The only feature that consistently distinguishes Zambesi Bantu is the root final vowel ų which has not been found in Equatorial Bantu. The loss of this vowel regularizes a phonologically deviant verb shape and might have occurred several times independently. At least, I find this more plausible than assuming the form -yúg- to be the retention.

The proposed "Wok" innovation is -ɔŋ 'hair', replacing the old nyuele , which is -juɪ́dɪ̄ (cl.11) in the Bantu reconstruction by Meeussen [1980]; the initial nasal is at least for Bantu analysable as the class 10 prefix which is the regular plural for class 11. Forms corresponding to -oŋ (a "second degree aperture" vowel is more appropriate for Bantu) seem to be missing in Zambesi Bantu. However, it is not at all clear what the general Bantu form should look like; the clue could come from Londo (A.11) ɲ-ungá if this item is cognate. On the other hand, it seems that the form -juɪ́dɪ́ has survived in several Equatorial Bantu languages, though the exact sound correspondences have not been worked out.[2] I therefore hesitate to accept this isogloss—be it lexical or phonological—as evidence against the internal unity of Bantu.

Finally, Bennett and Sterk suggest that "Wok" languages are distinguished from "Ungwa" languages by reflexes of an item baŋ 'red'. Reflexes of this root do indeed occur in Equatorial Bantu, e.g. Bafia (A.53) -ɓaŋ 'become red/ripe/soft'. However, while 'red' is not one of the most stable words in Bantu, reflexes of *-pɪ́- 'become burnt/cooked/hot/ripe/red' (with derived nouns and adjectives meaning 'fire', 'burnt grass', 'garden', 'hot', 'new', and 'red') appear in Equatorial and in Zambesi Bantu languages. (This root

[2] An old Noho (A.32) vocabulary gives menjede 'hair'. Other possible reflexes are found in A.40 and A.60, e.g. Numand (A.46) tu-úɲ , Nukalong (A.67) tuúɲe . The reviewer of this paper has also pointed out that "some Zone A languages show both *ɔŋ and *jųidi as 'head-hair' and 'body-hair'."

has a wide distribution within Niger-Congo.)

Lexicostatistics can provide no more than a first hypothetical outline of a genetic classification. Conclusive evidence is hard to get from isoglosses, probably because we are unable to systematize in a useful way the facts of semantic change and language contact. The most promising approach to the complex problem of subclassifying Bantu and Bantoid languages appears to lie in the search for irreversible and characteristic sound shifts. This task still lies ahead. For the time being I know of no compelling evidence to deny the genetic unity of Bantu, which is moreover strongly supported by lexicostatistic inspection of the similarity matrix provided by Bennett and Sterk.

APPENDIX

Table 1: Language names, numbers, and codes

1.	GR	Grebo
2.	NE	Newole
3.	AS	Asante
4.	LA	Larteh
5.	LE	Lelemi
6.	GA	Ga
7.	EW	Ewe
8.	GW	Gwari
9.	GD	Gade
10.	NU	Nupe
11.	IA	Igbira
12.	ID	Idoma
13.	IO	Igbo
14.	IG	Igala
15.	IF	Ife
16.	YO	Yoruba
17.	OR	Ora
18.	BI	Bini
19.	UR	Urhobo
20.	IS	Isoko
21.	DE	Degema
22.	IJ	Ijo
23.	AB	Abua
24.	EF	Efik
25.	OG	Ogoni
26.	KJ	Kaje
27.	CH	Chori
28.	AF	Afusare
29.	AT	Aten
30.	KM	Kambari
31.	JU	Jukun
32.	KP	Kpan
33.	TV	Tiv
34.	MB	Mambila
35.	EL	Eloyi
36.	TN	Tunen
37.	JA	Jarawa
38.	NY	Nyanja
39.	BO	Bobangi
40.	KK	Kikuyu
41.	KW	Kwanyama
42.	FU	Fula
43.	DY	Dyola
44.	TM	Temne
45.	MO	Mossi
46.	KS	Kassena
47.	MP	Mamprusi
48.	TL	Tula
49.	GB	Gbaya
50.	ND	Ndogo

Figure 1: BA Subclassification

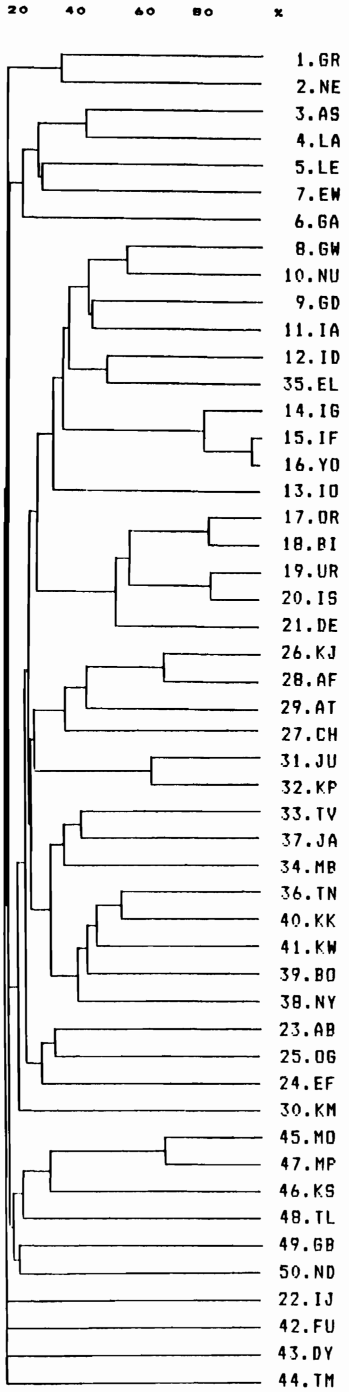


Figure 2: NN Subclassification

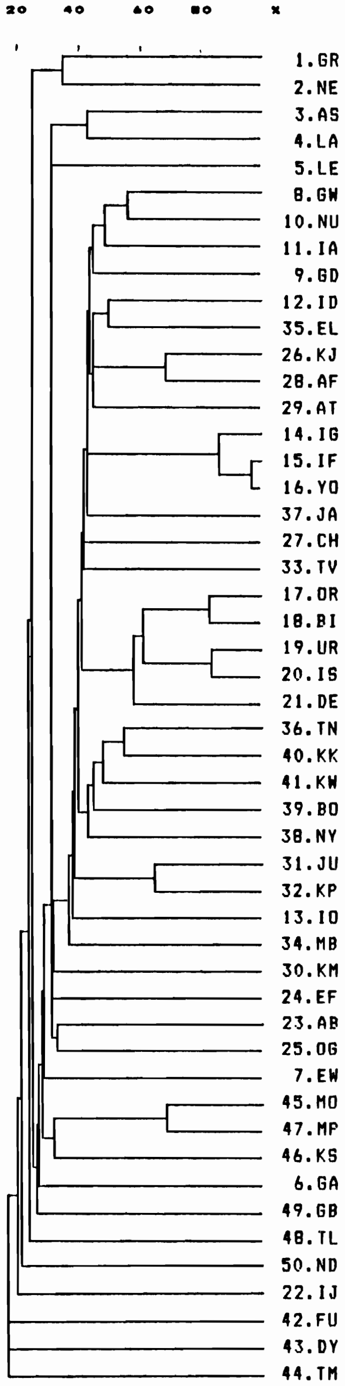


Figure 3: FN Subclassification

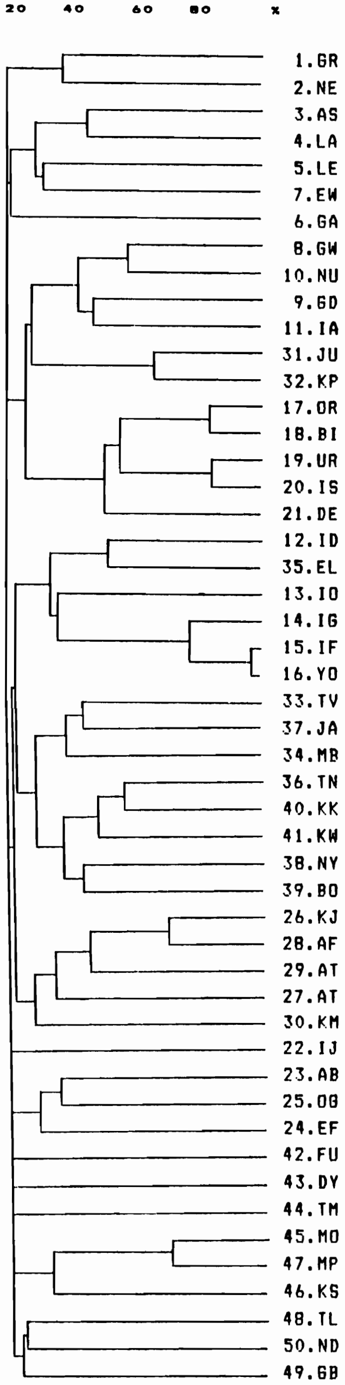


Table 2: NN, FN, and BA Cluster Analysis

	Nearest Neighbour		Furthest Neighbour		Branch Average	
	lg.x:lg.y	1/1000	lg.x:lg.y	1/1000	lg.x:lg.y	1/1000
cluster #1:	15 : 16;	970	15 : 16;	970	15 : 16;	970
cluster #2:	14 : 15;	860	19 : 20;	840	19 : 20;	840
cluster #3:	19 : 20;	840	17 : 18;	830	17 : 18;	830
cluster #4:	17 : 18;	830	14 : 15;	770	14 : 15;	815
cluster #5:	26 : 28;	690	26 : 28;	690	26 : 28;	690
cluster #6:	45 : 47;	690	45 : 47;	690	45 : 47;	690
cluster #7:	31 : 32;	650	31 : 32;	650	31 : 32;	650
cluster #8:	17 : 19;	610	8 : 10;	560	17 : 19;	573
cluster #9:	17 : 21;	580	36 : 40;	550	8 : 10;	560
cluster #10:	8 : 10;	560	17 : 19;	540	36 : 40;	550
cluster #11:	36 : 40;	550	12 : 35;	500	17 : 21;	530
cluster #12:	12 : 35;	500	17 : 21;	490	12 : 35;	500
cluster #13:	8 : 11;	490	36 : 41;	460	36 : 41;	470
cluster #14:	36 : 41;	480	9 : 11;	450	9 : 11;	450
cluster #15:	8 : 9;	450	3 : 4;	430	26 : 29;	440
cluster #16:	12 : 26;	450	26 : 29;	430	36 : 39;	438
cluster #17:	12 : 29;	450	33 : 37;	420	8 : 9;	435
cluster #18:	36 : 39;	450	38 : 39;	410	3 : 4;	430
cluster #19:	8 : 12;	440	8 : 9;	400	33 : 37;	420
cluster #20:	3 : 4;	430	33 : 34;	360	36 : 38;	408
cluster #21:	8 : 14;	430	1 : 2;	350	8 : 12;	378
cluster #22:	8 : 37;	430	36 : 38;	350	26 : 27;	370
cluster #23:	36 : 38;	430	13 : 14;	340	33 : 34;	365
cluster #24:	8 : 27;	420	23 : 25;	330	8 : 14;	359
cluster #25:	8 : 33;	420	26 : 27;	320	1 : 2;	350
cluster #26:	8 : 17;	410	12 : 13;	310	23 : 25;	330
cluster #27:	8 : 36;	400	45 : 46;	300	8 : 13;	328
cluster #28:	8 : 31;	390	5 : 7;	290	33 : 36;	317
cluster #29:	8 : 13;	380	3 : 5;	260	45 : 46;	310
cluster #30:	8 : 34;	370	23 : 24;	260	5 : 7;	290
cluster #31:	1 : 2;	350	33 : 36;	260	23 : 24;	285
cluster #32:	23 : 25;	330	26 : 30;	250	8 : 17;	278
cluster #33:	8 : 30;	320	8 : 31;	250	3 : 5;	275
cluster #34:	45 : 46;	320	8 : 17;	230	26 : 31;	269
cluster #35:	3 : 5;	310	48 : 50;	210	26 : 33;	257
cluster #36:	3 : 8;	310	12 : 33;	200	8 : 26;	247
cluster #37:	3 : 24;	310	48 : 49;	200	8 : 23;	237
cluster #38:	3 : 23;	310	12 : 26;	190	3 : 6;	225
cluster #39:	3 : 7;	290	3 : 6;	180	45 : 48;	225
cluster #40:	3 : 45;	280	1 : 3;	170	8 : 30;	214
cluster #41:	3 : 6;	270	1 : 8;	170	49 : 50;	210
cluster #42:	3 : 49;	260	1 : 12;	170	45 : 49;	195
cluster #43:	1 : 3;	250	1 : 22;	170	3 : 8;	183
cluster #44:	1 : 48;	240	1 : 23;	170	3 : 45;	181
cluster #45:	1 : 50;	210	1 : 42;	170	1 : 3;	175
cluster #46:	1 : 22;	200	1 : 43;	170	1 : 22;	171
cluster #47:	1 : 42;	170	1 : 44;	170	1 : 42;	170
cluster #48:	1 : 43;	170	1 : 45;	170	1 : 43;	170
cluster #49:	1 : 44;	170	1 : 48;	170	1 : 44;	170

Table 3: Similarity Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	GR	NE	AS	LA	LE	GA	EW	GW	GD	NU	IA	ID	IO	IG	IF	YO	OR	BI	UR	IS	DE	IJ	AB	EF	OG	KJ	CH	AF	AT	KM	JU	KP	TV	MB	EL	TN	JA	NY	BO	KK	KW	FU	DY	TM	MO	KS	MP	TL	GB	ND
1.GR	-																																																	
2.NE	35	-																																																
3.AS	17	17	-																																															
4.LA	17	17	43	-																																														
5.LE	17	17	31	26	-																																													
6.GA	17	17	25	27	20	-																																												
7.EW	17	20	26	27	29	18	-																																											
8.GW	18	17	19	17	23	17	25	-																																										
9.GD	18	17	26	21	23	17	23	43	-																																									
10.NU	21	17	20	17	27	18	20	56	42	-																																								
11.IA	25	24	23	18	27	20	25	40	45	49	-																																							
12.ID	21	20	27	22	30	24	29	41	43	43	44	-																																						
13.IO	17	17	24	17	26	17	26	23	28	27	30	34	-																																					
14.IG	24	17	24	18	24	17	27	34	36	33	32	38	34	-																																				
15.IF	22	17	21	17	25	17	25	29	34	35	37	37	37	77	-																																			
16.YO	24	17	26	18	27	17	23	34	38	40	38	43	38	86	97	-																																		
17.OR	18	20	23	17	24	17	29	29	30	28	28	35	25	33	35	40	-																																	
18.BI	18	20	20	17	23	17	25	26	26	25	27	31	26	32	34	41	83	-																																
19.UR	21	18	23	17	20	17	26	26	25	25	27	30	23	29	34	37	59	55	-																															
20.IS	20	17	22	17	20	17	29	27	27	27	29	31	23	30	34	38	61	54	84	-																														
21.DE	18	20	22	17	21	17	27	23	25	24	29	29	25	32	32	35	58	49	52	53	-																													
22.IJ	17	17	17	17	17	17	17	17	17	17	17	19	17	19	17	19	20	17	17	17	17	-																												
23.AB	17	17	18	17	21	17	17	18	17	20	22	26	29	18	18	19	22	19	21	17	25	19	-																											
24.EF	17	19	17	17	19	17	17	19	21	24	27	31	24	23	23	22	25	26	24	23	27	17	26	-																										
25.OG	17	17	20	17	20	17	20	19	20	20	23	22	25	26	23	26	20	23	21	19	26	17	33	31	-																									
26.KJ	17	17	20	17	23	17	23	22	27	24	25	29	25	32	31	35	24	22	21	21	25	17	20	28	23	-																								
27.CH	17	17	19	17	20	17	18	20	25	21	25	27	22	28	29	30	20	20	19	19	18	17	18	26	21	42	-																							
28.AF	17	17	17	17	21	17	20	23	27	26	28	32	25	30	29	32	25	23	21	21	25	17	21	29	23	69	42	-																						
29.AT	17	17	19	17	18	17	21	20	23	21	24	25	22	23	22	24	23	21	21	22	23	17	22	25	26	45	32	43	-																					
30.KM	17	17	17	22	17	18	17	22	22	25	24	21	29	22	19	19	20	20	20	21	21	17	17	22	19	29	26	30	25	-																				
31.JU	17	17	17	17	18	17	20	25	28	26	25	30	19	31	26	31	28	28	23	24	24	17	17	25	20	27	27	28	18	17	-																			
32.KP	19	17	17	17	24	17	24	32	32	30	29	39	24	37	33	39	26	30	23	24	24	20	19	28	23	34	29	34	23	17	65	-																		
33.TV	19	17	17	18	20	17	17	23	25	27	29	28	29	36	32	31	28	24	21	22	25	17	17	23	28	30	29	32	24	21	30	35	-																	
34.MB	17	19	19	17	17	17	17	24	25	24	25	25	20	28	26	27	20	21	21	21	24	17	17	21	20	25	25	25	25	24	22	24	37	-																
35.EL	21	23	27	24	31	17	25	30	38	31	32	50	31	35	32	37	32	30	25	28	31	18	25	31	22	45	35	45	35	24	34	32	31	24	-															
36.TN	20	22	22	26	20	17	18	23	31	28	31	28	29	31	31	28	27	24	22	23	25	17	22	29	27	32	32	29	29	32	28	30	37	34	38	-														
37.JA	17	17	21	20	26	17	26	23	25	27	27	31	28	32	30	33	27	26	24	26	26	17	18	26	23	29	25	28	26	19	21	18	42	36	43	38	-													
38.NY	17	17	17	17	21	17	17	23	27	24	24	21	23	28	25	28	23	23	21	21	25	17	18	21	22	27	27	25	22	22	21	24	37	26	31	35	36	-												
39.BO	17	17	21	21	20	17	20	19	23	20	23	28	23	26	23	24	23	22	21	21	26	17	22	25	26	32	31	32	23	22	25	30	32	31	38	40	33	41	-											
40.KK	18	19	21	22	19	17	20	20	26	25	29	25	29	28	28	40	24	22	24	25	28	17	22	26	26	31	32	29	24	32	24	28	37	31	34	55	33	43	45	-										
41.KW	17	17	18	18	17	17	19	17	21	18	21	25	21	26	25	26	22	21	22	21	25	17	17	22	23	27	27	25	24	29	23	25	36	26	30	48	34	42	45	46	-									
42.FU	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	-								
43.DY	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	-							
44.TM	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	-						
45.MO	17	18	28	23	25	18	18	17	23	18	21	23	19	23	22	25	24	22	18	17	22	17	17	19	17	21	19	20	17	17	19	23	22	24	25	22	24	22	23	17	21	17	17	17	-					
46.KS	19	19	24	19	21	18	17	21	26	21	23	25	18	18	17	18	20	19	21	20	23	17	19	17	17	19	21	19	21	17	17	17	24	20	25	23	21	23	20	21	23	17	17	17	32	-				
47.MP	17	17	23	20	23	17	17	17	21	17	18	24	17	22	21	22	22	22	18	18	22	17	17	21	17	17	17	17	17	17	17	17	22	21	24	23	24	20	23	17	21	17	17	17	69	30	-			
48.TL	17	17	18	17	20	17	20	17	18	17	17	17	17	17	17	17	17	17	17	17	17	17	21	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	24	24	18	-		
49.GB	17	17	20	17	20	17	21	17	23	18	23	22	18	17	23	23	18	18	21	21	20	17	17	17	17	17	17	17	17	18	17	17	17	17	18	26	20	17	18	18	21	17	17	17	18	21	20	20	-	
50.ND	18	17	17	17	17	17	18	17	17	17	18	19	21	21	21	20	20	20	20	17	20	17	18	18	21	17	17	17	18	17	19	17	20	20	17	21	17	17	21	21	19	17	17	17	17	17	17	21	21	-

REFERENCES

Bennett, Patrick R. and Jan P. Sterk. 1977. "South Central Niger-Congo: a reclassification." *Studies in African Linguistics* 8:241-273.

Dyen, Isidore, A. T. James and J. W. L. Cole. 1967. "Language divergence and estimated word retention rate." *Language* 43:150-171.

Guthrie, Malcolm. 1967-1971. *Comparative Bantu*, 4 vols. Farnborough: Gregg.

Meeussen, A. E. 1980. *Bantu Lexical Reconstructions*. Archives d'Anthropologie, 27. Tervuren: Musée Royal de l'Afrique Centrale.