Research in African Languages and Linguistics, 1(1): 1–10 (1990)

A LEXICOSTATISTIC CLASSIFICATION OF ỊJỌ DIALECTS

J. D. Lee (Loughborough) and Kay Williamson *(Port Harcourt)*

1. Previous classifications of Ịjọ

The Ịjọ dialects are spoken in the Niger Delta and in nearby riverine areas within the Rivers, Bendel and Ondo States of Nigeria.

Ịjọ (anglicized as Ijaw) is often referred to as a single language, but there is neither mutual intelligibility between all the dialects nor an accepted standard variety for the whole area.

Fig. 1 summarizes four earlier classifications. It is arranged for maximum comparability. Talbot's 1926 classification is 'tribal' (his term) rather than purely linguistic, since he includes three non-Ịjọ languages. Wolff (1959) based his classification on wordlists of Mein, West Tarakiri, Kolokuma, Nembe, Akassa, Kalabari (from Buguma), and Okrika, plus further information on dialects from his informants. Williamson 1965 is based on a lexicostatistic comparison (using the Swadesh 200 word list) of Kalabari, Nembe, Bụmọ, Kolokunma, and Kabou, plus wordlists of other dialects. Williamson 1972 (unpublished) is chiefly based on phonological and lexical innovations. Hansford et al. 1976 use this classification in a simplified form; in Fig. 1 features of the 1972 classification which do not appear in the 1976 one are parenthesized.

Fig. 1. Previous classification of Ịjọ

Talbot 1926	*Wolff 1959*	*Williamson 1965*	*Williamson 1972* *Hansford 1976*
		I Eastern	I <Eastern)
A. Kalabari with Okrikan	A. Kalabari with Okrika	A. South-Eastern	A. (North)-Eastern
		1. Kalabari	1. a. Kalabari
		2. Okrika	b. Okrika
		3. Bonny (Ibani)	c. Ibani
			2. Nkoro
B. Lower Ijaw	B. Brass-Nembe	B. Brass-Nembe	B. (South-Eastern)
1. Brass-Nembe	with Akassa	1. Brass-Nembe	1. Nembe
		2. Akassa	2. Akassa
2. Ogbinya		II Central	(II. Central)
[Non-Ịjọ]			C. (General) Izon
			(1. True Central)
3. Brass-Ijaw	Lower Ịjọ	C. South-Central	a. South-Central
			i. Apoi
		1. Bassan	ii. Bassan
		2. Olodiama	iii. Olodiama East
		3. Oporoma	iv. Oporoma
		4. Boma	v. Boma
			vi. Oiakiri
			vii. Mein
			viii. Tarakiri East
			ix. Ikibiri

Talbot 1926	*Wolff 1959*	*Williamson 1965*	*Nilliamson 1972* *Hansford 1976*
C. Western Ijaw 1. Warri	C. Upper Ijọ including Kolokuma	D. North-Central 1. Ekpetiama 2. Kolokuma 3. Gbanran	b. North-Central i. Ekpetiama ii. Kolokuma iii. Gbanran
	D. Western Ijo	E. North-Western	2. North-Western
		1. Ikibiri-Tarakiri E.	
		2. Ogboin	
	1. Tarakiri Kumbowei Kabowei	3: Tarakiri West 4. Kumbo 5. Kabo	a. Tarakiri West b. Kumbo c. Kabo
	2. Mein Seimbri	6. Mein Seimbri	d. Mein e. Seimbri
	3. Tuomo Operemor Beni	Tuomo Operemor	f. Tuomo g. Operemor
		F. South-Western	3. South-Western
	4. Iduwini Ogula	1. Eduwuni 2. Ogula	a. Iduwini b Ogulagha
2. Atissa [Non-Ijọ]		3. Oporoza 4. Arogbo	c. Oporoza d. Arogbo
3. Mini [Non-Ijọ]			e. Egbema f. Olodiama g. Furupagha
(Talbot's map places Mini in *Lower* Ijaw)		G. North-Eastern 1. Amegi (Biseni) 2. Okordia	D. (North-East-Central) a. Biseni b. Okordia

The present exercise was undertaken as the first stage of a lexicostatistic classification of all the languages of the Niger Delta. The list used was the Swadesh list as revised at Ibadan for use with African languages: the meanings *this, that, who, what, not, all, many, louse, bark, liver, stand, rain, cloud, burn, green, yellow, round,* have been replaced by *you/ye, three, four, five, child, fowl (chicken), goat, housefly, navel, roast, swallow, blow* (of *wind), steal, rope, saliva, give birth, bury.* Some words have been modified: *flesh* to *meat, grease* to *fat, foot* to *leg, swim* to *bathe, fly* to *jump, lie* to *lie down, sit* to sit down, say to *say (something), earth* to *ground, ash* to *ashes, path* to *road, mountain* to mountain/hill.

2. The lexicostatistic classification

Observations have been made on a shared common vocabulary consisting of the equivalents of each of 100 words in 33 different Ijọ dialects, and the table of data is complete. These dialects can be classified into one or more homogeneous groups by applying lexicostatistic procedures. The classification involves two stages:

(a) Conversion of the raw data into a similarity matrx, giving a measure of the similarity between each pair of dialects. The coefficients increase as a pair of dialects become more similar.

(b) Sorting the dialects into groups, on the basis of the similarity coefficients. Hierarchical classification has been used, where similar dialects are combined to form a cluster, which in turn can be combined into larger clusters.

Conversion of the data into a similarity matrix

The principle is essentially very simple. A pair of dialects are compared to find the total number of cognate and possibly cognate words which occur in the 100 word list. This implies 100 comparisons, but since up to three words are allowed which have the same meaning, the number of word comparisons may be as high as 100 x 3 x 3 = 900.

With 33 dialects, the first should be compared with each of the other 32 in turn. Then the second dialect should be compared with the remaining 31, and the third with the remaining 30. . . . and so on. This gives a total of 528 pairs which must be compared, hence the total number of comparisons needed may be as high as 528 x 900 = 475,200. . A computer programme was written to carry out this daunting number of comparisons. The maximum number of agreements is plainly 100, and these numbers were scaled into the range 0 to 1, and arranged as a triangular similarity matrix (Fig. 2).

Fig. 2, Listing of the Original Percentage Data as a Triangular Matrix

KAL	100															
OKR	96	100														
IBA	95	92	100													
NKO	84	84	82	100												
NEM	84	86	82	79	100											
AKA	79	80	77	77	92	100										
BUM	74	76	73	74	82	83	100									
ETA	74	74	73	74	78	78	95	100								
AGU	75	76	74	75	81	81	97	99	100							
IKI	75	74	75	74	78	78	90	94	93	100						
EKT	74	73	74	74	77	74	84	87	89	91	100					
KOL	76	75	75	75	80	80	90	94	93	96	98	100				
GBA	71	70	71	70	75	75	86	89	90	93	95	99	100			
KAB	77	77	76	77	81	80	89	93	92	93	91	98	89	100		
KUM	76	76	75	75	80	79	87	90	89	91	85	94	87	95	100	
WTA	77	77	76	78	82	81	90	94	93	95	91	98	91	99	95	100
MEI	76	76	75	76	82	82	89	92	92	92	88	94	89	94	91	94
TUO	77	77	76	77	82	82	89	92	92	93	89	95	90	96	93	96
OPE	74	74	73	74	78	78	84	88	87	89	87	91	88	91	89	91
QGB	77	76	76	77	88	80	92	96	95	98	90	97	91	95	92	96
OIY	74	75	73	73	80	81	94	95	94	93	89	94	91	92	90	92
OPO	75	76	74	75	82	82	98	98	99	93	89	93	89	92	90	93
IKE	70	71	69	69	79	80	88	86	90	82	84	87	83	84	83	84
KOR	71	72	70	70	80	81	91	86	89	81	82	86	82	82	81	83
OND	71	70	70	71	79	81	91	85	87	84	84	87	84	85	84	86
BAS	72	73	71	71	82	82	87	84	88	81	80	85	80	82	82	83
IDU	69	68	68	66	76	78	83 •	82	84	83	80	84	80	83	81	83
OGU	67	67	66	65	75	76	79	78	81	77	76	80	75	79	76	79
OPZ	69	69	69	69	77	78	82	82	85	82	82	85	80	84	83	84
ARO	70	70	69	70	78	79	86	84	86	83	83	87	81	85	82	85
ORU	77	77	75	76	79	73	79	80	84	80	81	82	79	81	79	82
OKD	71	72	71	72	74	71	78	79	83	80	84	84	81	81	78	82
BIS	67	66	66	64	68	66	71	70	73	74	75	75	73	76	72	76
	KAL	OKR	IBA	NKO	MEN	AKA	BUM	ETA	AGU	IKI	EKT	KOL	GBA	KAB	KUM	WTA

	MEI	TUO	OPE	OGB	OIY	OPO	IKE	KOR	OND	BAS	IDU	OGU	OPZ	ARO	ORU	OKD	BIS
MEI	100																
TUO	99	100															
OPE	94	97	100														
OGBO	94	95	90	100													
OIY	92	95	92	94	100												
OPO	90	91	87	95	94	100											
IKE	86	88	85	85	92	89	100										
KOR	85	86	82	84	90	89	95	100									
OND	87	88	84	86	90	89	94	95	100								
BAS	85	87	83	84	89	87	95	95	92	100							
IDU	85	87	86	84	88	84	91	87	86	92	100						
OGU	81	83	82	79	83	80	86	84	82	89	91	100					
OPZ	87	89	89	83	87	84	89	86	87	90	91	91	100				
ARO	88	89	88	85	90	86	93	90	90	93	94	90	94	100			
ORU	79	80	78	82	78	83	77	76	75	75	71	69	73	74	100		
OKD	80	81	80	83	79	81	76	75	77	74	72	69	76	74	82	100	
BIS	73	74	75	75	71	74	69	68	72	66	69	65	71	70	75	85	100

Sorting the dialects into groups

Hierarchical cluster analysis operates on a similarity matrix which contains coefficients of similarity between pairs of dialects. All the dialects are taken initially as separate clusters. Then the two dialects with the highest similarity are merged, and considered as a single cluster. The similarity between the new cluster and all of the others is then re-defined in one of a number of ways, which are described later. The next pair of clusters with the highest similarity are then merged, the new cluster is re-defined, and the process is repeated until all the dialects belong to a single cluster.

Four different methods of clustering have been used on the data. They allworkby agglomerating dialects into clusters, but they differ in the criteria used for re-defining the similarity of a new cluster:

1. Nearest Neighbour or Single Linkage cluster analysis,

In this method, the similarity between two clusters is the greatest similarity between any two dialects, one in each cluster.

2. Furthest Neighbour cluster analysis

This method calculates the similarity between two clusters as the least similarity between any two dialects, one in each cluster.

3. Average Linkage cluster analysis

The similarity coefficients of a newly formed cluster are calculated as the mean of the coefficients of the two parent clusters from which it was formed, with each of the other clusters.

4. Centroid cluster analysis

This method is very similar to the average linkage method, except that the similarity of a new cluster is re-calculated as the weighted mean of the similarities of the merged clusters with each of the remaining clusters. Thus the distance between two groups of languages depends on all of the languages, rather than just the extreme member of each group.

The single linkage (nearest neighbour) criterion is the fastest to compute, and usually emphasises the separateness of clusters. If there are no isolated clusters, a single large cluster (possibly with a few outliers) is produced. A possible disadvantage of this method is that it may produce "chines", that is chains of dialects, with each only a small distance from its immediate neighbours, but with the ends of the chain perhaps far apart.

The other methods usually maintain compact clusters. without regarding the possibility that two similar dialects may be assigned to different major clusters. If there are no isolated clusters, these methods will sub-divide the dialects in a somewhat arbitrary manner, which may change radically by the addition or deletion of a single dialect. The different clustering methods and their merits are discussed in Sneath and Sokal 1973, Jardine and Sibson 1971, Gower 1967.

Cluster analysis was performed on these data, using each of the four methods described. Calculations were performed using the program GENSTAT. Dendrograms produced by each of these four methods are shown in Figs. 3, 4, 5 and 6. The most striking feature of these is their close similarity.

Fig. 3. Nearest Neighbour or Single Linkage Cluster Analysis

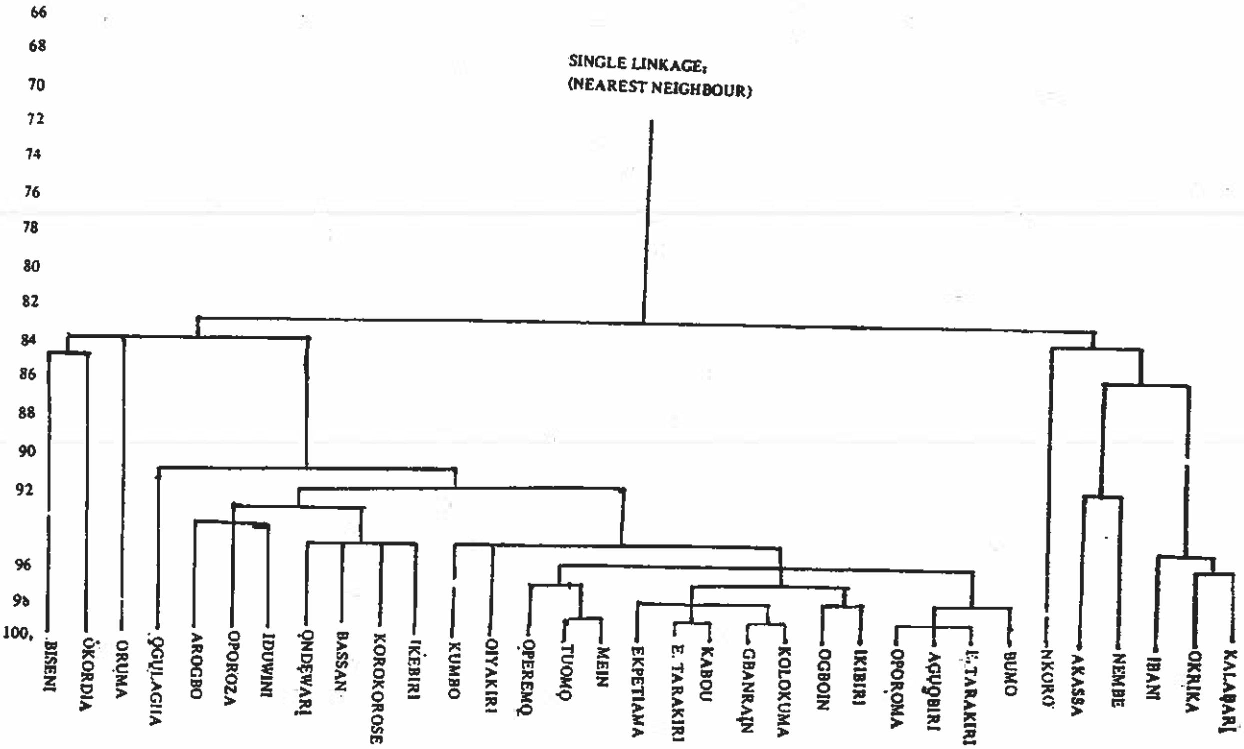


Fig. 4. Furthest Neighbour Cluster Analysis

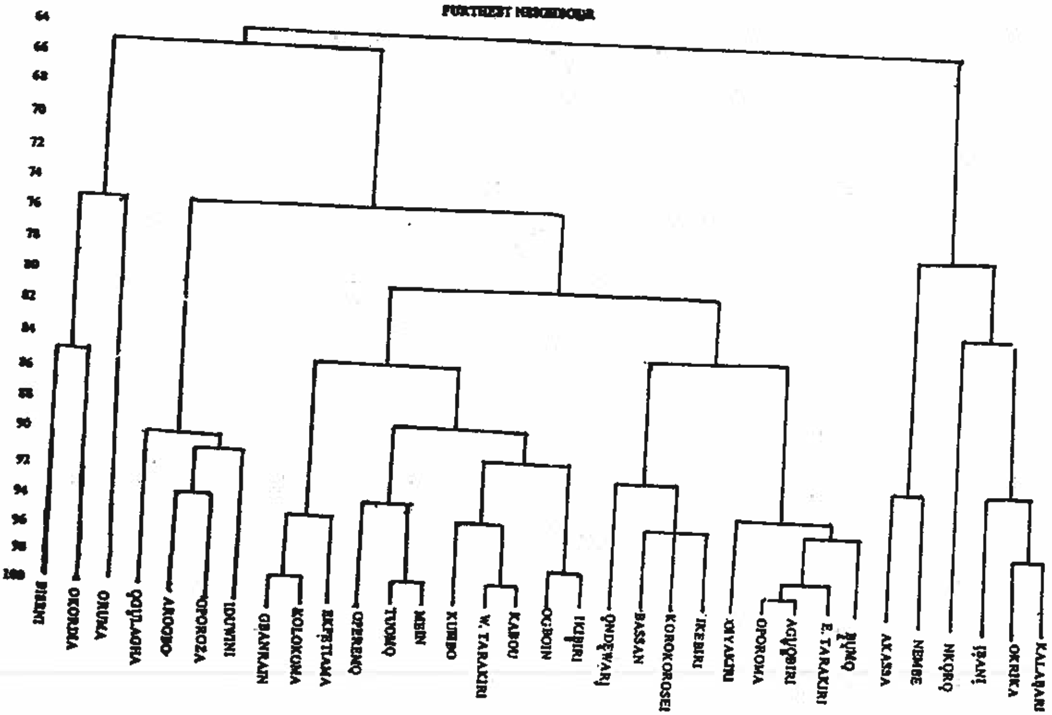


Fig. 5. Average Linkage Cluster Analysis

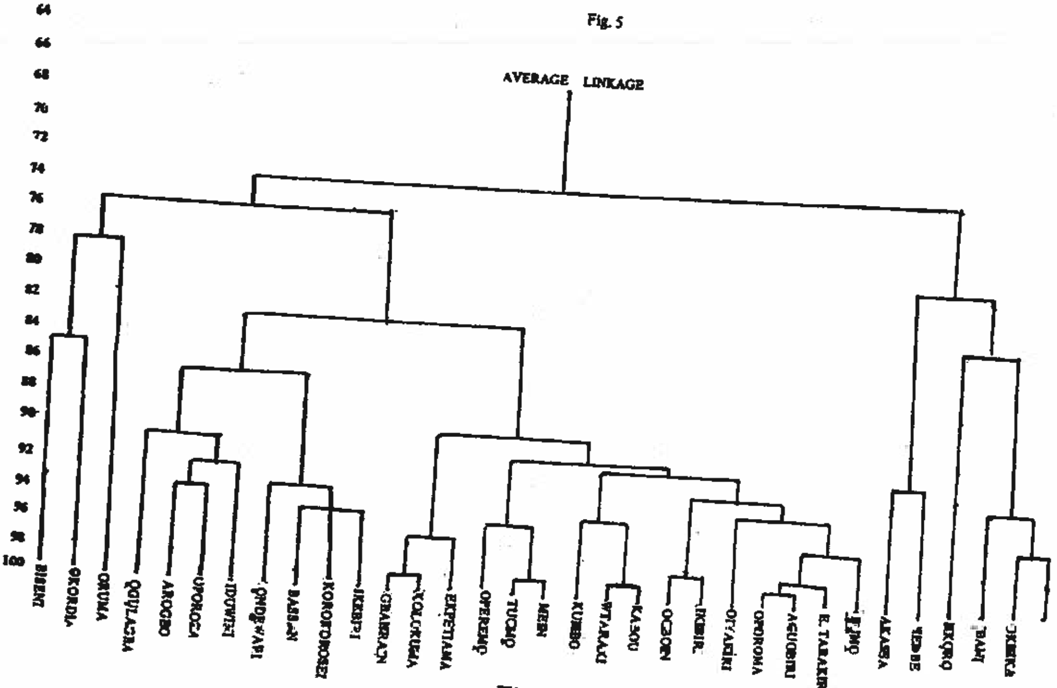
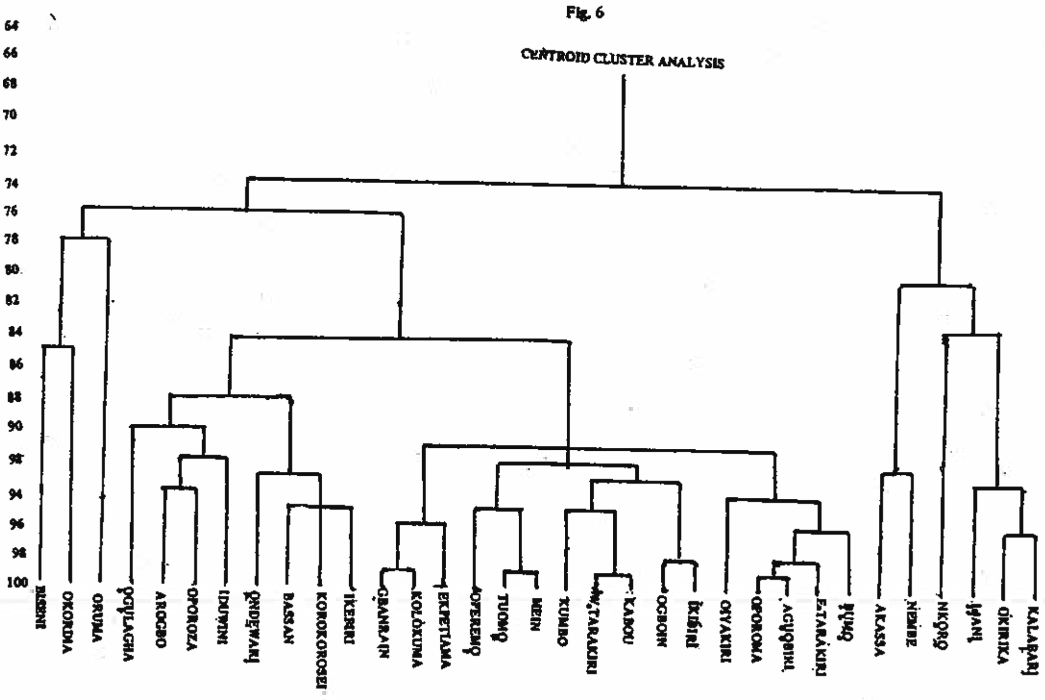


Fig. 5

Fig. 6. Centroid Cluster Analysis



3. Results

1. As would be expected, the fusion heights for nearest neighbour occur much lower than in the furthest neighbour method, and the average and centroid methods are intermediate.

2. The dialects at the right of the dendrograms KAL, OKR, IBA, NEM, AKA and NKO form a cluster of identical shape in the furthest neighbour, average and centroid methods, and in the nearest neighbour method this differs only in the point of fusion of NKO.

3. The dialects at the left hand side of the dendrograms: ORU, OKD and BIS form a cluster of identical shape in the furthest neighbour, average and centroid methods, and it differs in the nearest neighbour method only in that ORU fuses to the pair OKD and BIS at the same height as fusion with the central group.

4. The 24 dialects in the centre of the dendrogram from BUM to OGU form a closely-knit group. There are many close similarities, for example the fusion of BUM, ETA, AGU, OPO,, and of IKE, KOR, BAS, OND. There are some minor differences between the different methods of classification, but these are not considered significant.

The deletion of a single dialect may radically change the order of clustering, particularly when there are many fusions close together at similar heights. All four clustering methods (nearest neighbour, furthest neighbour, average and centroid) have been applied to the data with the OPO dialect omitted, and again to the full data set with just the ORU dialect omitted. These caused only very minor changes in the fusion heights, but did not change the shape of the tree in any way.

4. Discussion

The differences obtained from the various clustering methods will now be compared with the results of earlier classifications. First, the nearest neighbour method will be contrasted with the other three methods.

1. **Nkọrọ.** The nearest neighbour method classifies Nkọrọọ as an entity of the same order as Kalaḅarị-Okrika-Ịḅanị and Nembe-Akassa; this corresponds to Jenewari s (1977) classification. The other methods treat it as more closely linked to Kalaḅarị-Okrika-Ibani than to Nembe-Akassa; this corresponds to Williamson's (1972) classification. Both alternatives have, therefore, been previously proposed, and cannot be used to decide between the methods.

2. **Oruma.**'This dialect was first studied a few years ago, and was at once classified with Biseni and Okordia. This is in line with the other methods rather than the nearest neighbour one.

3. **Ọgụlagha.** The nearest neighbour method opposes Ọgụlagha (marginally) to all the other Central dialects, whereas the other methods group it with Iduwini, Oporoza and Arogbo, in aggreement with the classifications by wolff (1969) and Williamson (1965, 1972).

4. **Ekpetiama.** The nearest neighbour method separates Ekpetiama slightly from Kolokuma and Gbanrain whereas the other methods keep them in the same cluster, as in Williamson 1965, 1972.

In three out of four cases, therefore, the nearest neighbour method clearly gives results which are less in accordance with independent classifications than the alternative methods. This strongly suggests that the nearest neighbour method is the least suitable for lexicostatistic classifications.

We next compare the furthest neighbour method with the other two methods.

1. **Ikebiri-korokorosei-Bassan-Ọndẹwarị.** The furthest neighbour method links this cluster with the Ḅụmọ-East Tarakiri-Agụọbiri-Oporoma-Ọiyakiri cluster, in accordance with their common classification as South-Central dialects (Williamson1965, 1972). The other methods link it with the Iduwini-Oporoza-Arogbo-Ọgụlagha cluster and separate it from the other South-Central dialects. Thus the furthest neighbour method gives the best results.

2. **General.** The greater fusion heights for the furthest neighbour method make use of the lower percentages in the similarity matrix. In glottochronology, percentages are converted into time depths. Since the intention is to estimate the time since variation began in the protolanguage, not merely since the dialects became fully differentiated, the lower percentages, or greater fusion heights, appear to be more useful.

We therefore conclude that the Ịjọ data strongly suggests that the furthest neighbour method is the most appropriate for lexicostatistics when closely-related lects are under consideration.

5. **Nomenclature.** The major division between the dialects at the right of the dendrograms and the rest is confirmed by a lexical difference; those to the right use *gbọrị* as the qualifying word for 'one', while the rest use *kẹnị*. We therefore call those to the right the GBỌRỊ group and the rest the KẸNỊ group.

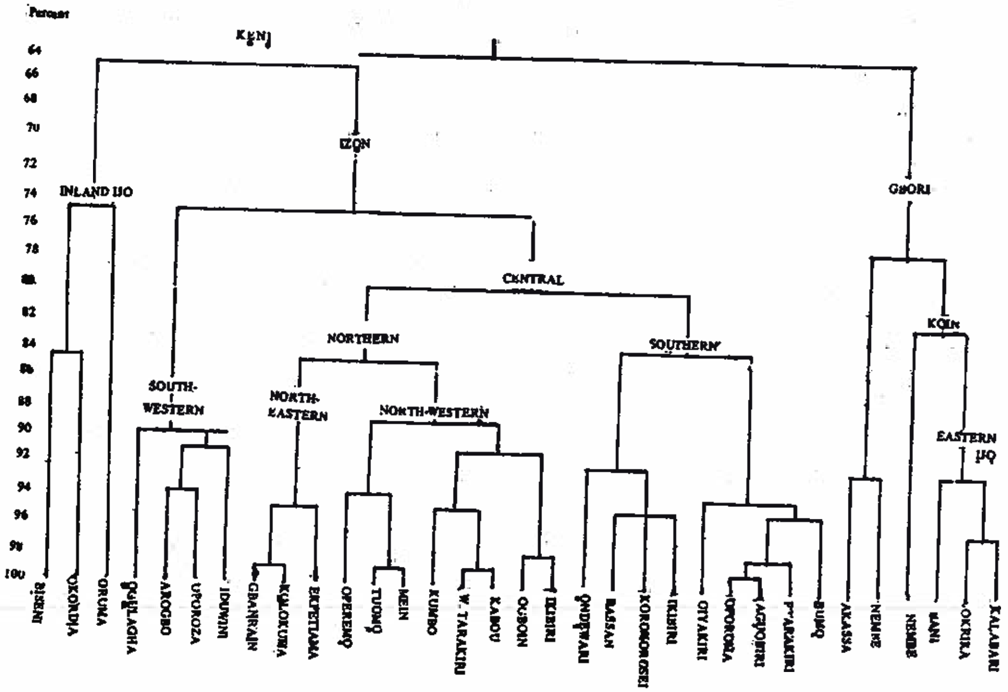
Within the Gbọrị group, recent work by Harry (1989) has confirmed the position of Nkọrọọ as intermediate between Nembe/Akassa and Kalaḅarị/Okrika/Ịḅanị, but somewhat nearer to the latter. We accept his proposal of the acronym KOIN for Kalaḅarị/Okrika/Ịḅanị/ Nkọrọọ and Jenewari's (personal communication) proposal of EASTERN ỊJỌ for Kalaḅarị/Okrika/Ịḅanị. Nembe or Nembe/Akassa is the partner group to Koin.

The Kẹnị group divides clearly into the dialects at the left of the dendrograms and the rest. Those at the left are all distant from the ocean and are therefore called INLAND ỊJỌ. The remaining dialects are called ỊZỌN, the speakers' usual name for themselves and their language.

Within Ịzọn, the major division is between Iduwini/Oporoza/Arogbo/Ọgụlagha and the rest. The former group are called SOUTH-WESTERN, in accordance with earlier terminology, and the latter CENTRAL. The Central dialects are divided into NORTHERN and SOUTHERN. The Northern dialects are divided into NORTH-EASTERN and NORTH-WESTERN. We have not thought it necessary

to provide labels for dialects which are differentiated at 90% or higher. .
The proposed classification and labels are presented in Fig. 7.

Fig. 7. Classification with Proposed Names for Groupings



6. Historical implications

Fig. 8 shows the dialects on a rough map. The major division between GBỌRỊ and KẸNỊ dialects corresponds to a great extent with a wedge of non-Ịjọ languages, particularly those of the Central Delta group, running deep into the Ịjọ area. It appears possible that the differentiation of the Ịjọ dialects into GBỌRỊ and KẸNỊ dialects began with the intrusion of Central Delta speakers into areas of the Delta earlier settled by Ịjọ speakers.

The INLAND IJỌ dialects are located adjacent to, and in the case of Oruma surrounded by, non-Ịjọ languages. It is suggested that their differentiation from the Ịzọn dialects was caused, slightly later than the first split, by a westward expansion of non-Ịjọ peoples which partly isolated them from their fellow-Ịjọ speakers and soon also from one another

Fig.8. Sketch-Map Showing Ịjọ Dialects

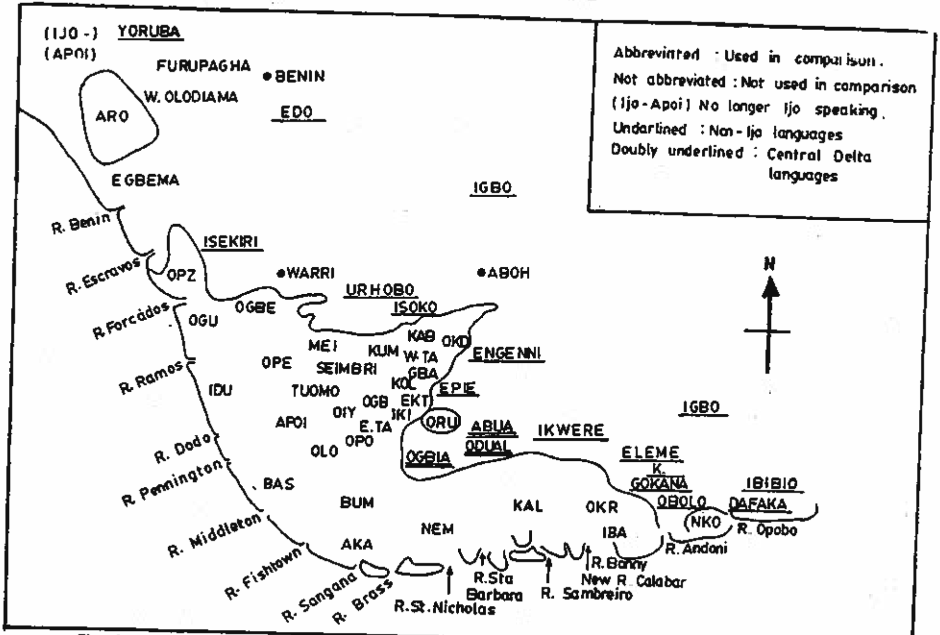


Fig. 8 : Sketch-map, showing Ijo dialects

Acknowledgements

This work was initiated with the aid of a Senate Research Grant to Kay Williamson from the University of Port-Harcourt.

The authors are grateful to the Computer Centres at the University of Port-Harcourt, the University of Loughborough, and the Manchester University Regional computer Centre, for the use of facilities.

An earlier version was presented at the colloquium on African Linguistics, Leiden, in 1982. We are grateful to colleagues there, especially Juliana Kuperus and Thilo Schadeberg, for their useful comments.

References

Gower, J.C. 1967. "A comparison of some methods of cluster analysis". *Biometrics* 23.626–637.

Hansford, K.J. Bendor-Samuel, and R. Stanford. 1976. "An index of Nigerian languages". *(Studies in Nigerian Languages,* 5.) Ghana:. Summer; Institute of Linguistics.

Harry, O.G. 1989. "A Comparative reconstruction of Proto-koin (Eastern Ịjọ and Nkọrọ) phonology". M.A. thesis. Port-Harcourt: University of Port Harcourt. Department of African Languages and Linguistics.

Jardine, N., and R. Sibson. 1971 *Mathematical taxonomy*. New York: Wiley.

Jenewari, C.E.W. 1977 *"Studies in Kalabari syntax." Ph.D. Thesis. Ibadan: University.*

Sneath, P.H., and R.R. Sokal. 1973 *Numerical taxonomy.* London and San Francisco: Freeman.

Talbot, P.A. 1926 (new impression 1969). *The peoples of Southern Nigeria.* London: Cass.

Westermann, D., and M.A. Bryan. 1952 (new ed. 1970). *Languages of West Africa.* London: Dawsons of Pall Mall for International African Institute.

Williamson, K. 1965. *A grammar of the Kolokuma dialect of Ịjọ.* West African Language Monographs. 2. Cambridge: University Press.

Williamson, K. 1972. The Language of the Ịjọ peoples Paper read at the seminar on the Peoples of South Eastern Nigeria. Nsukka Uniersity of Nigeria, (Mimeograph.)

Wolff, H. 1959. "Niger Delta languages I: classification." *Anthropological Linguistics* 1:8.32–53.