



Internal relationships in southern Bantu

C. H. Borland

To cite this article: C. H. Borland (1986) Internal relationships in southern Bantu, South African Journal of African Languages, 6:4, 139-141, DOI: [10.1080/02572117.1986.10586665](https://doi.org/10.1080/02572117.1986.10586665)

To link to this article: <https://doi.org/10.1080/02572117.1986.10586665>



Published online: 24 Oct 2012.



Submit your article to this journal [↗](#)



Article views: 11



View related articles [↗](#)

Internal relationships in southern Bantu

C.H. Borland

Department of African Languages, University of Cape Town, Rondebosch, 7700 Republic of South Africa
Present address: Department of Computer Science, University of Zimbabwe, Harare, Zimbabwe

Received September 1985

This article is concerned with the internal classification of the southern Bantu languages and more specifically with the relationships between Shona and these languages. It is shown that the relationship between Shona and Venda is pivotal to an understanding of this classification.

Hierdie artikel handel oor die interne klassifikasie van die suidelike Bantoetale en meer spesifiek die verwantskap tussen Shona en dié tale. Daar word bewys dat die verwantskap tussen Shona en Venda die spil vorm in die begrip van hierdie klassifikasie.

Shona is classified separately from the south-eastern languages in Ehret (1972). This early paper remains influential and this basic classification is retained in more recent work such as Ehret (1982). This original classification was made on the basis of percentage cognacy measured over a 90-word basic vocabulary sample described in Figure 1.

Shona					
55	Venda				
37	59	Sotho			
35	44	54	Nguni		
38	52	46	44	Chopi	
41	54	57	55	57	Tsonga

Figure 1 Percentage cognacy measured over a 90-word basic vocabulary example; Nguni represents the average of Xhosa and Swati (from Ehret, 1972:10)

In terms of these numerical estimates of relationship Venda is clearly more closely related to the other south-eastern languages than Shona is. Further, Ehret's detailed analysis of the patterns of cognacy that underlie these estimates shows that Venda shares eight of the ten south-eastern Bantu lexical innovations in this sample, which implies a long period of common ancestry. Shona cognates with the south-eastern languages, on the other hand, are mainly joint retentions from earlier proto-Bantu periods and make no contribution to later internal southern-Bantu developments. Similarly, Shona and Venda share only two lexical innovations over this sample, implying a shorter period of common ancestry. But if Venda is clearly south-eastern Bantu and Shona is not, the surprisingly close relationship between Venda and Shona is very anomalous.

Ehret argues that the closeness of this Venda – Shona relationship must be the result of borrowing. But using borrowing as an explanation of the level of relationship

measured over cognate vocabulary is not very satisfactory. The very definition of cognacy — vocabulary inherited from a common source item via regular sound shifts — is intended to exclude borrowing. Thus it is hardly surprising that Ehret is able to cite only one example of a Shona – Venda borrowing in his vocabulary sample. But if the presence of merely apparent or false cognates in a sample of formally cognate vocabulary is in principle undetectable, then the borrowing explanation becomes an essentially unfalsifiable hypothesis. Notice that this consideration affects any argument which involves cognacy — including the previous evidence of shared innovations, some proportion of which may well be mere apparent cognates. One must recall also that it was this very problem of indeterminacy which led to the original postulation by Swadesh (1952, 1955) of the basic vocabulary sample employed by Ehret in this analysis. The assumption was that concepts in this sample are in some informal sense universal and essential and therefore that the associated vocabulary items would be unlikely to be borrowed. This basic vocabulary list has been employed repeatedly by Ehret and others for this very reason. It is, therefore, a little surprising to note the easy acceptance of the indeterminacy of the results based on this sample.

An alternative approach to the classification of these languages is now considered, using larger vocabulary samples, more precise numerical estimates of relationship, and an explicit formal classificatory algorithm. The theoretical basis and advantages of this alternative approach over Ehret's rather informal methodology have been discussed elsewhere (Borland, 1984) and will not be repeated here.

The data for this alternative analysis are the comparative word lists of Guthrie (1967). These data have both advantages and disadvantages compared to Ehret's. An immediate disadvantage is that Chopi and Tsonga are not included in the sample of languages. On the other hand, data on over 2000 vocabulary items are available, from which a number of standard vocabulary samples may be drawn. Guthrie attempted to minimize the effect of false cognates on this data by including only those items with a certain minimum areal distribution, namely three of his language zones. As a direct result, data representing possible shared innovations with a narrower distribution are not available.

- In our analysis four subsets of this data were used:
- 100-word basic vocabulary sample of Swadesh (1955);
 - 200-word basic vocabulary sample of Swadesh (1952);
 - 550-word general vocabulary sample of Guthrie (1967);
 - 2235-word total vocabulary sample of Guthrie (1967).

The numerical estimate of relationship used is the taxonomic distance coefficient of Sneath & Sokal (1973:124). Unlike Ehret’s measure of cognacy, which is based only on joint retentions, this coefficient is sensitive to all possible patterns of joint retention and joint and disjoint replacement. The formal classificatory algorithm used is the unweighted pair group method with arithmetic averages (UPGMA) of Sneath & Sokal (1973:228 – 234).

Numerical estimates of relationship measured over these four vocabulary samples are described in Figures 2 – 5. The taxonomic distance coefficient varies from zero to one for complete dissimilarity to complete similarity respectively. The first three decimal places are given in the figures. Nguni represents the average of Xhosa and Zulu.

It is clear that these results are very similar to Ehret’s. In every case Venda is apparently more closely related (i.e. less distant) to the other south-eastern languages than Shona is, and Shona is apparently most closely related to Venda.

Shona			
385	Venda		
445	359	Sotho	
451	379	347	Nguni

Figure 2 Taxonomic distance coefficients for 100-word basic vocabulary sample

Shona			
430	Venda		
472	386	Sotho	
475	410	392	Nguni

Figure 3 Taxonomic distance coefficients for 200-word basic vocabulary sample

Shona			
565	Venda		
617	511	Sotho	
653	582	508	Nguni

Table 4 Taxonomic distance coefficients for 550-word general vocabulary sample

Dendrograms describing the UPGMA classifications based on the estimates of relationship are described in Figures 6 – 9. For comparison, Ehret’s own results

Shona			
475	Venda		
505	419	Sotho	
504	437	407	Nguni

Figure 5 Taxonomic distance coefficients for 2235-word total vocabulary sample

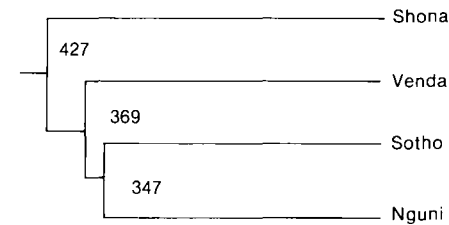


Figure 6 UPGMA classification of taxonomic distance coefficients over 100-word basic vocabulary sample

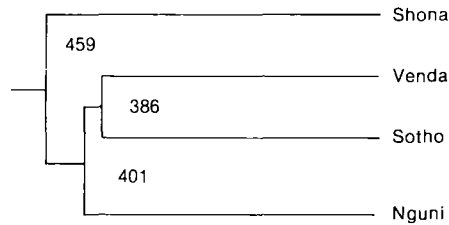


Figure 7 UPGMA classification of taxonomic distance coefficients over 200-word basic vocabulary sample

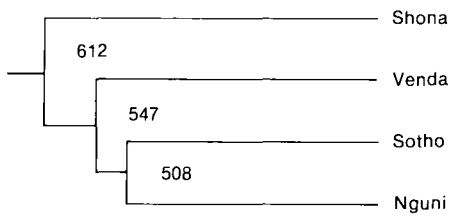


Figure 8 UPGMA classification of taxonomic distance coefficients over 550-word general vocabulary sample

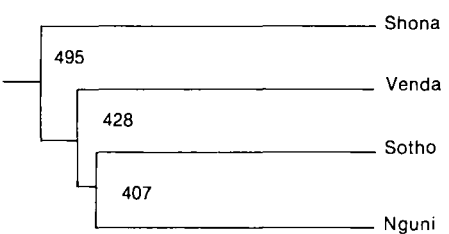


Figure 9 UPGMA classification of taxonomic distance coefficients over 2235-word total vocabulary sample

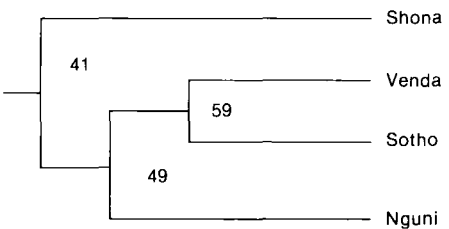


Figure 10 UPGMA classification of percentage cognacy over 90-word basic vocabulary sample (data from Figure 1)

(omitting Chopi and Tsonga) were processed using this algorithm to produce the dendrogram given in Figure 10. The values of the taxonomic distance coefficient (or, in the case of Figure 10, the percentage cognacy) at which clusters form, are provided.

Despite the similarity of the estimates which underlie these classifications there are significant differences between them, chiefly in the sequence in which languages join the nuclear cluster. Thus in Figures 6, 8, & 9; Sotho, Venda, and then Shona join the nuclear Nguni cluster in that sequence; while in Figures 7 & 10, Sotho and Venda are linked together before they join the nuclear cluster. In other respects the classifications are similar.

The question can now be addressed whether Shona should be classified as a separate sub-branch from the south-eastern languages. There are several approaches to this question, apart from Ehret's rather informal and intuitive one. For these purposes this question is best answered in terms of group heterogeneity, for which numerical estimates are available in our classification. It is clear from examination of the dendrograms that the differences between Shona and the remaining south-eastern languages are generally significantly smaller than the internal differences within these south-eastern languages. Interestingly, this difference is smallest in the formal classification of Ehret's own results described in Figure 10. On this basis then, it would be argued that Shona is best classified as south-eastern Bantu.

It should be noted that this analysis of the classifications of these south-eastern languages is extracted from a larger study, which includes 28 languages covering the whole Bantu area. In all four analyses of this larger language set, according to the different vocabulary samples described above, there is a clear break between this south-eastern subgroup (including Shona) and the other subgroups that are discernible by the group heterogeneity criterion. This increases confidence in the genetic reality of this subgrouping.

A final confirmation of this classification is the archaeological evidence of Huffman (1978). Although this is

certainly incomplete in this area and subject to a variety of methodological interpretations (Hall, 1983), the basic proposition of an Early Iron Age Shona origin south of the Limpopo, with a Later Iron Age presence in Zimbabwe, still seems reasonable. Linguistic and archaeological counter-evidence against Ehret & Kinsman's (1981) proposal for an Early Iron Age Shona presence in Zimbabwe itself has already been presented in Borland (1984; in press).

References

- Borland, C.H. 1984. Conflicting methodologies of Shona dialect classification. *S. Afr. J. Afr. Lang.*, Vol. 4 No. 1.
- Borland, C.H. (in press). *Problems in the linguistic reconstruction of Zimbabwean prehistory*. Centre for African Studies, Cape Town: University of Cape Town.
- Ehret, C. *et al.* 1972. Outlining southern African prehistory: a reevaluation AD 100 – AD 1500. *Ufahamu*, Vol. 3.
- Ehret, C. 1982. The first spread of food production to southern Africa. In Ehret, C. & Posnansky, M. (eds). *Archaeological and linguistic construction of African prehistory*. Berkeley: University of California Press.
- Ehret, C. & Kinsman, M. 1981. Shona dialect classification and its implications for Iron Age history in southern Africa. *Int. J. Afr. Hist. Stud.*, Vol. 14 No. 3.
- Guthrie, M. 1967. *Comparative Bantu*. Farnborough: Gregg Press.
- Hall, M. 1983. Tribes, traditions and numbers: the Americal model in southern African Iron Age ceramic studies. *S. Afr. Archaeol. Bull.*, Vol. 38.
- Huffman, T.N. 1978. The origins of Leopard's Kopje: an 11th century Difaquane. *Arnoldia*, Vol. 8(23).
- Sneath, P. & Sokal, R. 1973. *Numerical Taxonomy*. San Francisco: Freeman.
- Swadesh, M. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proc. Am. Philos. Soc.*, Vol. 96.
- Swadesh, M. 1955. Towards greater accuracy in lexicostatistic dating. *Int. J. Am. Linguist.*, Vol. 21.