

---

**The linguistic importance of  
language isolates: the African case**

Jean-Marie Hombert & Gérard Philippson

---

Proceedings of Conference on  
**Language Documentation & Linguistic Theory 2**

Edited by Peter K. Austin, Oliver Bond, Monik Charette,  
David Nathan & Peter Sells

13-14 November 2009 School of Oriental and African Studies, University of London

Hans Rausing Endangered Languages Project  
Department of Linguistics  
School of Oriental and African Studies  
Thornhaugh Street, Russell Square  
London WC1H 0XG  
United Kingdom

Department of Linguistics:  
Tel: +44-20-7898-4640  
Fax: +44-20-7898-4679  
[linguistics@soas.ac.uk](mailto:linguistics@soas.ac.uk)  
<http://www.soas.ac.uk/academics/departments/linguistics>

Hans Rausing Endangered Languages Project:  
Tel: +44-20-7898-4578  
Fax: +44-20-7898-4349  
[elap@soas.ac.uk](mailto:elap@soas.ac.uk)  
<http://www.hrelp.org>

© 2009 Jean-Marie Hombert & Gérard Philippson

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, on any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author(s) of that part of the publication, except as permitted by UK copyright law.

ISBN: 978-0-7286-0392-9

This publication can be cited as:

Jean-Marie Hombert & Gérard Philippson. 2009. The linguistic importance of language isolates: the African case. In Peter K. Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 2*. London: SOAS.

or:

Jean-Marie Hombert & Gérard Philippson. 2009. The linguistic importance of language isolates: the African case. In Peter K. Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 2*. London: SOAS. [www.hrelp.org/eprints/ldlt2\\_15.pdf](http://www.hrelp.org/eprints/ldlt2_15.pdf)

**The linguistic importance of language isolates: the African case**  
 JEAN-MARIE HOMBERT<sup>1</sup>. & GÉRARD. PHILIPPSON<sup>2</sup>  
*Dynamique du Langage (CNRS, Université de Lyon)<sup>1</sup> &*  
*Institut National des Langues et Civilisations Orientales (Paris) / Dynamique du*  
*Langage (CNRS, Université de Lyon)<sup>2</sup>*

1. ENDANGERED LANGUAGES

Endangered languages have become a major object of investigation over the last decade, due to the realization that linguistic, as well as biological, diversity is increasingly threatened by global societal and ecological changes. It is however clear that a complete description of even the most threatened languages – which number in the hundreds – is, unfortunately, well beyond the means of the scientific community. Defining clear criteria for the choice of priorities is crucial, both for justification of the investment of research time and for funding decisions. To take the case of one African country, Gabon, half of the 50-odd languages have less than 1000 speakers (Hombert, 2009):

**Table 1**  
 Languages of Gabon

<10 speakers	<100 speakers	<1000 speakers	<10,000 speakers	>10,000 speakers
Batsi	Gevia	Apinji	Aduma	Fang
Irimba ??	Ivili	Baka	Akele	Kota
Yongwe??	Metombolo	Bakaningi	Bakwele	Myene
	Mwesa	Bongwe	Benga	Nzebi
	Ngubi	Koya	Getsogo	Obamba
	Tumbidi	Latsitsege	Gevove	Punu
	Kande	Ndambomo	Gisir	Sangu
		Ndasa	Lumbu	Teke
		Seki	Mahongwe	
		Shamayi	Mbaouin	
		Shiwa	Ndumu	
		Sigu	Shake	
		Simba	Ungom	
		Tsengi	Varama	
		Vungu	Vili	
			Wanzi	
			Wumbu	

Drawing from our African experience, we argue that endangered language studies should assign a high priority to the study of linguistic isolates - that is, languages

with either questionable or unknown genetic affiliations. Specifically, the investigation of linguistic isolates yields invaluable information for improving the general classification of the world's languages, for enriching knowledge of specific typological traits, and for elucidating early population movements (in particular, in the case of 'hunter-gatherer' languages being absorbed by languages spoken by invading agriculturalists).

Specifically, the place of the African continent as the home of *Homo sapiens* is generally not in dispute among paleontologists and geneticists. Human genetic diversity in Africa is much greater than anywhere else and the oldest remains of *homo sapiens* have been found there. On that account, it seems rather paradoxical that African linguistic diversity appears much lower than that in, say, New Guinea, whose oldest settlement surely does not predate c. 60,000 BP, or the Americas (30,000 BP ?). One should thus investigate the reasons for this apparent uniformity.

## 2. AFRICAN LANGUAGE CLASSIFICATION

A first answer which comes to mind is that linguistic diversity on the African continent has been underestimated by existing classifications. Ever since Greenberg (1963), the Africanist scientific community has by and large accepted his classification of African languages into four large phyla<sup>1</sup>: Niger-Congo (aka Congo-Kordofan or Niger-Kordofan), Nilo-Saharan, Afro-Asiatic and Khoisan. This classification was admittedly far from revolutionary in its broad outlines, since the Khoisan phylum was more or less identical with the click language group recognized at least since Bleek (1929), Afro-Asiatic was equivalent to the old Hamito-Semitic family, to which Greenberg boldly added not only the 'Chado-Hamitic' languages of Westermann (long felt to be related to Hamito-Semitic, e.g. by the French semiticist M. Cohen), but also the Chadic ('non-Hamitic') languages that Westermann wanted to keep separate from the others on account of their lack of grammatical gender opposition. The other two phyla 'Niger-Congo' and 'Nilo-Saharan' were largely based on Westermann's West- and Ostsudansprachen, with a few modifications, like moving the Songhai group (Mali, Niger) into Nilo-Saharan, but most notably the provocative inclusion of Bantu languages as a sub-branch of the Benue-Congo family within Niger-Congo (although Greenberg himself acknowledged that Westermann tacitly supported this interpretation, while the French traditional Africanist school - Homburger, Delafosse, etc. - considered Bantu and 'Sudanic' languages as the two branches of a 'Negro-African' phylum).

There does not seem currently to be any doubt about the inclusion of Bantu in Niger-Congo, nor indeed about the cohesion of the latter, albeit with a great many disagreements about its internal structure. The unity of the other three phyla

---

<sup>1</sup> In spite of his methodology being submitted to severe criticism from comparative linguists (Dixon, Campbell, etc.)

however is still disputed: least of all Afro-Asiatic, where it is only the place of Omotic (South-Western Ethiopia), or indeed its inclusion within the phylum, that is still open to queries - at the very least for some of the component languages. But most Khoisan specialists do not now regard all click languages as being related: the inclusion of Hadza (Tanzania) is generally rejected, whereas of the other sub-groups the only one to form an accepted genealogical grouping is Khoe (which admittedly contains the largest number of languages), all other languages being potential isolates.

But the most disputable phylum appears to be Nilo-Saharan : even setting apart the case of Songhai, whose inclusion is rejected by many specialists, it is notable that the final elaboration of the phylum by Greenberg took place fairly late. In his original classification, published in the early 1950's, he only recognized an 'Eastern Sudanic' group, included along with 'Central Sudanic' and two smaller language groups, Kunama and Berta, within a larger 'Macro Sudanic' or 'Chari-Nile' phylum. Only in his 1963 book was the decision taken to join 'Macro Sudanic' with 'Central Saharan' (Kanuri, Teda, Zaghawa), Songhai and three smaller groups of the Ethiopia-Sudan region (Maban, Fur and Koman) which had been left isolated in the first version.

One of the most telling proofs of the problematic status of 'Nilo-Saharan' is that two recent attempts at reconstructing 'proto-Nilo-Saharan' (Bender 1996 and Ehret 2001) end up with two very different - in fact incompatible - internal classifications of the phylum. Even 'Eastern Sudanic', which should presumably prove most resistant to restructuring, does not escape entirely unscathed: Greenberg's 'Teuso' (nowadays more generally called Kuliak, a remnant language group in eastern Uganda) is moved by Bender outside of the 'Eastern Sudanic' family altogether, whereas Ehret firmly retains it (in fact many contemporary researchers would consider Kuliak an isolate).

It is thus quite likely that a number of languages have been misclassified, and indeed there is a growing feeling among specialists of various linguistic areas in Africa that about 12 to 15 languages should be classified as isolates. Our own feeling on the matter is that this figure might well be too low. Whatever the case, and even with this proviso in mind, there is still much less heterogeneity in Africa than in New Guinea (where no less than 60 different phyla have been identified - some of them with an admittedly very small membership - for a total of 800-odd languages).

Another cause for this loss of linguistic diversity should probably be sought in the expansion of food-producers (agriculturalists and/or pastoralists) into territories formerly inhabited exclusively by hunter-gatherers. The expansion of Bantu languages into Central, Eastern and Southern Africa constitutes a particularly telling example. Processes of state-building and other centralized or semi-centralized polities certainly played a role in more recent times, as exemplified by Vansina (1990) for the equatorial forest and Schoenbrun for the Great Lakes area (1998), among others. It can be surmised that expansion of chiefly power into ever-widening regions implied the concomitant growth of the language of the court at the expense of the languages of subject peoples.

**Table 2**  
Language isolates

<b>Languages</b>	<b>Country</b>	<b>Linguistic affiliation</b>
≠Hoan	Botswana	Used to be considered language isolate. Is now considered part of Khoe-Kwadi group
Aasax or Aramanik	Tanzania	Probably South Cushitic; contains non-Cushitic lexicon
Bangi-me	Mali	Strongly influenced by Dogon
Bung	Cameroon	Probably Adamawan
Gomba	Ethiopia	
Hadza	Tanzania	Isolate
Imeraguen	Mauritania	Hassaniyya (Arabic) restructured on an Azêr base (Soninke)
Irimba	Gabon	Non-Bantu lexicon
Jalaa or Centuum	Nigeria	Heavy Adamawa influence
Kara	Central African Republic	
Kujarge	Sudan, Chad	possible Chadic affiliation
Kwadi	Angola	Used to be considered language isolate. Is now considered part of Khoe-Kwadi group
Laal or Gori	Chad	Chadic substrate. Adamawa?
Lufu	Nigeria	
Luo	Cameroon	Extinct?
Mawa	Nigeria	
Meroitic	Sudan	Probably North Eastern Sudanic family
Mpre or Mpra	Ghana	Probably not Niger-Congo
Oblo	Cameroon	Unclear position within Adamawa. Extinct?
Ongota or Birale	Ethiopia	Possibly Afro-Asiatic (Cushitic? Omotic?) Now speak Ts'amakko.
Oropom	Uganda	Extinct?
Pre	Ivory Coast	Probably Niger-Congo
Rer Bare	Ethiopia	They now speak Somali
Sandawe	Tanzania	Probably linked to Khoe-Kwadi group
Shabo	Ethiopia	Nilo-Saharan?
Weyto	Ethiopia	Eastern Sudanic or Cushitic?
Wutana	Nigeria	
Yeni	Cameroon	Extinct

### 3. IDENTIFYING AFRICAN LANGUAGE ISOLATES

Table 2 summarizes the information on the status of those African languages which have been considered linguistic isolates.

### 4. UNCLASSIFIED LANGUAGES AND SUBSTRATA

When we talk of linguistic ‘isolates’ we should look beyond the 12 to 15 languages which seem to remain unaffiliated and consider also languages whose broad characteristics allow their inclusion in an identified family or group without assigning to them a precise place in the genealogical tree- which might of course be partly due to defective information<sup>2</sup>. We might mention here the cases of Dogon (Niger-Congo), Dahalo (Cushitic), Sandawe (Khoisan), Songhai and Kuliak (if their Nilo-Saharan affiliation is accepted) etc. Some of these languages exhibit features of exogenous origin which might prove extremely significant for an interpretation of their history. A notable example might be Dahalo, an undoubtedly Cushitic remnant language of coastal Kenya<sup>3</sup>, which has in its lexicon almost 100 lexical items containing clicks - these being of course unknown in other Cushitic languages and unrelated to other click languages of Eastern and Southern Africa. One should also keep in mind in this respect, the particular lexicon identified by Bahuchet in various Pygmy languages.

This in turn brings us to the whole question of substratum lexicon in African languages. Whereas in general *within* well-defined families, the amount of cognate vocabulary is important in the realm of basic terms - but not necessarily of cultural items, as even the example of Bantu languages shows - *between* families undoubtedly belonging to the same phylum, rates of cognation in basic vocabulary are very low. One might mention within Afro-Asiatic, the case of the various Cushitic subgroups in comparison with one another (Eastern vs. Northern, Central and even Southern) or even more strikingly the case of Berber with the rest of the phylum: whereas morphological features suffice to put beyond doubt the inclusion of Berber within the northern or ‘Boreo-afasian’ subdivision of Afro-Asiatic, the amount of shared vocabulary with the rest of the phylum is amazingly small<sup>4</sup>. One could hypothesize that some of the idiosyncratic lexicon is

---

<sup>2</sup> And indeed, out of 2,000-odd African languages, how many can be said to be classified on the basis of satisfactory data? Most of them are only known from short word lists, and if one excepts the 500 or so Bantu languages whose typological characteristics and genealogical closeness are normally sufficient to allow to classify them with some certainty, even from a word list, there remain several hundreds of languages (the Chadic group is a good example) whose classification must be taken on faith alone.

<sup>3</sup> Classified as South Cushitic by Ehret (1974 and *passim*) but more likely to constitute an independent branch of Eastern Cushitic.

<sup>4</sup> To the extent that Ehret (1995) excluded Berber altogether from his very liberal reconstruction of Afro-Asiatic lexicon.

due to some substratum element (Capsian?) present in Northern Africa before the advent of Berber speakers (around the 5th millennium B.C.) - a blending of the two populations - and hence their languages - would tally with the fact that there doesn't seem to be any great discontinuity in lithic industries between epipaleolithic and neolithic times (Brett & Fentress 1996). The same situation probably obtains in many if not most African language families and a careful collating of lexical material should be able to shed much light on our subject.

## 5. CONCLUSION

It can be seen that by reevaluating Greenberg's comprehensive but over-synthetic classification and taking into consideration putative language isolates in the light of the previous discussion, we should be able to achieve a more elaborate list of some twenty different groups (major phyla + language isolates) instead of only four, which would put African language classification in a much less peculiar position in comparison with the rest of the world.

## REFERENCES

- Bender, Lionel M. 1996. *The Nilo-Saharan languages: A comparative essay*. München -Newcastle: Lincom Europa.
- Bleek, Dorothea F. 1929. *Comparative vocabularies of Bushman languages*. Cambridge: Cambridge University Press.
- Blench, Roger. 2007. Endangered languages in West Africa. *Language diversity endangered*. In by Matthias Brenzinger (ed.), 140–62. Berlin, Germany: Mouton de Gruyter.
- Brett, Michael & Elizabeth Fentress. 1996. *The Berbers*. Malden & Oxford: Blackwell.
- Connell, Bruce. 2007. Endangered languages in Central Africa. *Language diversity endangered*. In by Matthias Brenzinger (ed.), 163–78. Berlin, Germany: Mouton de Gruyter.
- Dimmendaal, Gerrit J. 2008. Language ecology and genetic diversity on the African continent. In *New Directions in Historical Linguistics*, Lyon, May 12–14, 2008.  
<http://www.ddl.ishlyon.cnrs.fr/colloques/index.asp?Action=Edit&Langue=FR&Page=NDHL>
- Dimmendaal, Gerrit and Erhard Voeltz. 2007. Endangered languages of Africa. In Christopher Moseley (ed.), *Encyclopaedia of the world's endangered languages*, pp. 579-634. London: Routledge.
- Ehret, Christopher. 1974. *Ethiopians and East Africans*. Nairobi: East African Publishing House.
- Ehret, Christopher. 1995. *Reconstructing Proto Afroasiatic (Proto Afrasian)*. Berkeley/Los Angeles/London: University of California Press.

- Ehret, Christopher. 2001. *A Historical-Comparative reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe Verlag.
- Greenberg, Joseph H. 1963. The Languages of Africa. *International Journal of American Linguistics*, 29(1).
- Hombert, Jean-Marie. 2009. La diversité culturelle de l'Afrique est menacée. *La Recherche* 429, 36-39.
- Mous, Maarten. 2003. Loss of Linguistic Diversity in Africa. In Mark Janse & Sijmen Tol (eds.) *Language death and language maintenance: theoretical, practical and descriptive approaches*, 157-170. Amsterdam: John Benjamins.
- Sands, Bonny 2009. Africa's Linguistic diversity. *Language and Linguistics Compass* 3/2, 559–580.
- Schoenbrun, David L. 1998. *A Green Place, a Good Place*. Oxford: James Currey.
- Vansina, Jan. 1990. *Paths in the Rainforests*. Oxford: James Currey.