

ISTVÁN FODOR

**La classification des langues négro-africaines
et la théorie de J. H. Greenberg***

I. — IMPERFECTIONS DU SYSTÈME DE CLASSIFICATION
DE GREENBERG

La classification des langues africaines ne peut pas être considérée, même de nos jours, comme un problème résolu. La cause en est, d'une part, le fait qu'on n'a pas encore achevé la description synchronique des centaines de langues et parlars africains par des moyens modernes (à l'exception de quelques douzaines de langues bien connues), et que nous n'avons que des notations trop brèves et scientifiquement insuffisantes sur la majorité de ces langues, et, d'autre part, qu'on ne possède pas de documents linguistiques, sauf pour quelques langues, faute de monuments écrits. Il s'ensuit que la possibilité de rapprochements synchroniques modernes et de comparaisons typologiques se limite à un petit nombre de langues seulement et qu'en outre il n'y a qu'une possibilité très restreinte d'emploi de la méthode comparative et historique classique, en dehors des langues bantu.

Toutefois, une division tripartite traditionnelle s'est formée au cours du siècle passé, qui fut acceptée jusqu'à ces dernières années : langues soudano-guinéennes, famille bantu et langues des Bochimans et Hottentots ou khoisanides¹. Parmi ces trois familles, seules les langues bantu présentent une unité généalogique incontestable (démontrée par les travaux de Meinhof), tandis que les langues du Soudan et de la Guinée ne forment qu'un conglomerat géographique des langues différant les unes des autres, dont quelques groupes seulement, par exemple les langues atlantiques occidentales (*Westatlantische*

* Cet article est la récapitulation succincte — avec une documentation réduite — de l'étude intitulée "The Problems in the Classification of the African Languages" (Center for Afro-Asian Research of the Hungarian Academy of Sciences, Studies on Developing Countries, n° 5, Budapest, 1966), parue en 1965 en hongrois dans le journal *Nyelvtudományi Közlemények* (Publications Linguistiques).

1. Selon VOEGELIN (1964 : 9-10) l'autonomie du groupe des langues nilotiques et nilo-chamitiques et le problème de leur généalogie ont rendu indispensable la distinction d'un quatrième groupe (sans compter les langues chamitiques qui sont traitées dans un cinquième groupe avec d'autres langues africaines).

Sprachen), peuvent être considérées comme apparentées grâce à l'activité de Westermann.

En dehors de propositions pour la classification de quelques groupes de langues africaines, aucune conception proposant des changements radicaux n'est apparue dans la première moitié de notre siècle. Mais, entre 1949 et 1954, une transformation révolutionnaire est survenue sous l'influence des articles de Joseph H. Greenberg publiés dans le *Southwestern Journal of Anthropology*. Ces articles ont eu un grand retentissement dans le monde entier et l'opinion est généralement répandue parmi les africanistes que la théorie de Greenberg a offert une nouvelle perspective à cette spécialité scientifique. L'auteur a republié son étude — en ayant modifié quelques thèses sous l'influence des critiques et l'ayant complétée de nouveaux résultats — dans un volume spécial de *l'International Journal of American Linguistics* (Part II) en 1963 (réédition en 1966). Dans le présent article je me réfère à l'édition de 1963.

Greenberg voit clairement la faiblesse des classifications précédentes, et c'est pourquoi il veut construire sa théorie sur la seule base des critères linguistiques, en excluant les critères ethnologiques et géographiques. Il fait une « comparaison en masse » (*mass comparison*) dont le principe est que les comparaisons des unités linguistiques ne sont valables qu'à condition expresse de ressemblance conjointe des sons et de la signification, refusant de se contenter de la conformité exclusive soit des sons, soit de la signification. En outre, il tient compte non seulement des racines, c'est-à-dire des unités du vocabulaire, mais aussi des morphèmes grammaticaux, c'est-à-dire qu'il exige la force probante de la grammaire. Enfin, ce qui est le plus important, il ne tire pas ses conclusions seulement de quelques cas pris au hasard, mais du témoignage d'une foule de données.

Malheureusement une exigence très importante fait défaut à ces principes, ce qui — comme nous le verrons plus tard — ruine la valeur de la méthode entière de Greenberg, tout comme la négligence de ce même critère avait déjà également égaré beaucoup de ses prédécesseurs. Ce critère indispensable, c'est l'existence des lois phonétiques, qui est la pierre de touche de toute comparaison génétique. C'est à l'aide des lois phonétiques que Meinhof a prouvé définitivement l'origine commune des langues bantu et en a inféré les éléments du bantu commun, alors que, à cause de la négligence complète des lois phonétiques, il n'a pas réussi à prouver l'origine chamitique du masai. Quant aux données de Greenberg, rassemblées par lui-même ou citées d'après des travaux d'autres chercheurs, ce sont les comparaisons du peul et du sérère (1963:26-27), de l'efik (fi) et du bantu commun (1963:33-34), des langues kordofaniennes (1963:149-150), et des langues nilotiques (1957) qui révèlent des lois phonétiques, indépendamment même de l'intention de l'auteur.

Quelle est la valeur de la documentation du chercheur américain ? Dans quelle mesure sa démonstration est-elle probante ? Malheureusement, toute l'accumulation des données ne démontre rien sauf dans les cas où des lois phonétiques subsistent. Les classifications et les comparaisons de Greenberg peuvent être acceptées tout au plus comme une hypothèse de travail, et de nouvelles recherches sont nécessaires pour examiner chacun de ses thèmes, pour les prouver ou les rejeter.

Je vais présenter des tentatives de comparaisons où la correspondance des phonèmes des langues en question est dépourvue même d'une apparence de régularité. Voyons quelques exemples concernant les comparaisons de mots.

Dès l'introduction, qui traite des questions méthodologiques (1963:4), il y a quelques unités lexicales dont les correspondances sont indiquées pour huit langues africaines appartenant à deux familles : 1. le berti (siga), 3. le teda (toubou), 6. le kanuri et 8. le zaghawa appartiennent au groupe *nilo-saharien* ;

2. le kotopo, 4. l'ahlō (ago), 5. le bantu commun et 7. l'efik (fi) au groupe nigéro-congolais. Le teda et le kanuri présentent, entre les sept unités de mot, deux correspondances qui répondent à des lois phonétiques :

	« un »	« deux »	« trois »	« main »	« œil »	« oreille »	« bouche »
Teda	<i>toro</i>	<i>ču</i>	<i>agozo</i>	<i>daho</i>	<i>samo</i>	<i>sumo</i>	<i>či</i>
Kanuri	<i>tilo</i>	<i>ndi</i>	<i>yasko</i>	<i>kela</i>	<i>sim</i>	<i>sumo</i>	<i>či</i>

S'il n'y avait pas davantage de données à notre disposition, on pourrait risquer l'hypothèse que, les phonèmes /s/ initial et /m/ à l'intérieur du mot correspondent régulièrement dans ces deux langues (voir les mots des unités « œil » et « oreille »). Par contre, les deux mots de l'unité « bouche » sont inutilement conformes l'un à l'autre, son par son, parce qu'il n'y a pas d'exemples sûrs pour contrôler les correspondances du phonème /č/. Les deux mots pour « deux », *ču* et *ndi*, peuvent être en corrélation étymologique de la même manière (dans ce cas ce serait justement l'exemple contraire de la correspondance précédente), mais on n'en a aucune preuve, Greenberg même ne les considérant pas comme connexes. Il est possible que les mots pour « un » soient d'origine commune, car ils en ont l'air, mais rien de précis n'est connu sur l'alternance de /t/ initial et de /r~l/ à l'intérieur du mot.

Mais une contradiction décisive se trouve dans les couples de mots suivants où Greenberg croit découvrir une identité étymologique totale :

	« un »	« deux »	« trois »	« main »	« œil »	« oreille »	« bouche »
Berti	<i>sang</i>	<i>su</i>	<i>soti</i>	<i>yung</i>	<i>sing</i>		<i>a</i>
Teda	<i>toro</i>	<i>ču</i>	<i>agozo</i>	<i>daho</i>	<i>samo</i>	<i>sumo</i>	<i>či</i>

Si les mots de l'unité « deux » sont réguliers, ceux de l'unité « œil » ne le sont pas, parce que dans ce cas la forme du teda devrait être *čamo ; par contre, si la correspondance pour « œil » est régulière, celle pour « deux » ne l'est pas, car dans ce cas la forme du teda devrait être *su.

Les exemples du tableau I sont pris dans la documentation comparative du chapitre sur les langues nigéro-congolaises (1963:13-24). Les numéros d'ordre des unités de mots et des groupes de langues sont ceux indiqués par Greenberg. Les données expressément contradictoires du point de vue de la loi phonétique sont signalées par des caractères différents¹.

Je souligne ainsi la contradiction la plus importante :

Les formes du bantu commun impliquent la différence de deux phonèmes séparés /p — b/. Mais si les données du mossi, pour les unités 4. et 49. /p — p/, sont régulières, alors le mumuyé les contredit /b — p/.

1. Il faut remarquer, pourtant, que je n'ai pu dresser de tableau que de manière intuitive et qu'il n'est pas certain qu'il soit apte à illustrer ce qui est indiqué en tête (occlusives bilabiales initiales) ou qu'il reflète les variantes de la réalisation du même son.

TABLEAU I. — LISTE DES MOTS COMPARÉS POUR LES LANGUES

UNITÉS	GROUPES DE					
	1		2		3	
4 - « Ventre »	Kisi Banyun Cobiana Cassanga	<i>puli</i> <i>be:r</i> <i>abu</i>	Bambara Susu (1) Kono	<i>furu</i> <i>furi</i> <i>ebu</i>	Mossi Dagomba Tobote	<i>pu(ga)</i> <i>puri</i> <i>(de)po(l)</i>
7 - « Noir »	Temné (2)	<i>bi</i>	Bozo-Soninké Dyula Kweni (3)	<i>bibi</i> <i>fi</i> <i>beri</i>	Kulango Kabré (2) Doghosié	<i>bi</i> <i>biyo</i> <i>biri(ge)</i>
9 - « Sein »	Biafada Mandyak Konyagi	<i>(a)bili</i> <i>pil</i> <i>(i)bal</i>			Tobote Dagari Gurma	<i>(de)bil</i> <i>bere</i> <i>(li)be(li)</i>
10 - « Enfant »	Peul Serer-Sin Balante	<i>bɔi</i> <i>bi</i> <i>(m)bi</i>			Mossi Siti Bariba	<i>bi(ga)</i> <i>bi</i> <i>bii</i>
49 - « Blanc »	Dyola Limba Gola	<i>fur</i> <i>fufu</i> <i>fua</i>	Mendé Sya Boko	<i>puru</i> <i>foro</i> <i>pura</i>	Awuna Sisala (7) Mossi	<i>pōa</i> <i>pula</i> <i>pel(γa)</i>

Cet exemple pourrait être complété par bien d'autres.

Greenberg a établi ses preuves non seulement sur la concordance des mots, mais aussi sur celles des éléments grammaticaux. Je voudrais mentionner quelques exemples où les concordances grammaticales sont tout à fait contradictoires avec les conformités de vocabulaire, au point de vue des lois phonétiques.

Pour vérifier la parenté des langues du Soudan oriental et du groupe chari-notique, l'auteur présente (1963:117) les correspondances du morphème verbal

ADAMAWA-ORIENTALES : OCCLUSIVES BILABIALES INITIALES

LANGUES		4	5	6	
Likpe Idoma Igala Ijo	(ka)fo ipu efu furo	Bute Janji Proto-Bantu	bur ɛpuru *pu	Mumuyé Taram Nielim Nzakara Bwaka Barambo	buru fu hul vuru bu bulu
Ahlō Kukuruku (2) Idoma (4)	bibi bi obi	Reshe (Yawri) Boritsū Efik (2)	ubiru ebr obubit	Jen Daka Mumuyé Zandé Nzakara Gbanziri Ndogo Sango (3) Banda (5)	bi vir(ki) viri bi bibiri bi bvibvi bi bi
Likpé Akposo Nupé	(lɛ)mi (ɛ)vi (e)be	Kentu Akunakuna Proto-Bantu	(a)mɛ ebi *(de)-bede	Messo (Fali) Yasing Jen Mumuyé	bet(ku) bi mi mi
Adelé Ewe Gbari	ebi vi (e)bi	Koro Kamuku Proto-Bantu (6)	ebiya boo *biad-	Mangbei Yasing Koké Gbaya Ndogo	bi(ga) bie ba be bvi
Adelé Ewe (7) Twi	fu fu fufu	Efik Bute Tiv	fua, fia eburi pupu	Mumuyé Mbum (7) Vere Ngbandi	puru fu bule vulu

- (1) « abdomen ».
(2) « être noir ».
(3) « nuit ».
(4) « obscurité ».

- (5) « nuitamment ».
(6) « accoucher ».
(7) « être blanc ».

négatif *m* ou *b* : 1. nubien *m-*, par exemple *esk* « être capable » ; *mesk* « être incapable » ; 2. didinga *ma* « prohibitif » ; 3. barea (baria) *ma* ; 7. mérarit *m-* ; 8. dagu (dadjo) du Darfour *ba* ; 9. shilluk *ba*. Dans les matériaux de comparaison lexicologique du groupe du Soudan oriental (1963:96) la rubrique 12. « être doux » contient les données suivantes : 2. didinga *med* ; dinka *mit* ; shilluk *met*. C'est

dire qu'en vertu des concordances grammaticales, le mot *shilluk* devrait avoir la forme **bet*, ou alors la conformité des données du *didinga* n'est pas régulière. Mais si nous acceptons que la concordance des mots est régulière, alors, au contraire, la forme du *shilluk* devrait être **ma* au lieu de *ba*.

Il arrive plus d'une fois que les correspondances des éléments grammaticaux se contredisent aussi. Greenberg (1963:131) dénombre les affixes contenant un phonème /*k*/ de l'accusatif et du pluriel (points 19. et 25.) comme un des arguments parlant en faveur de la parenté des langues nilo-sahariennes. Un exemple du /*k*/ pluriel est le suffixe *-go* du *maba* (ouadaïen) dans des pronoms personnels, l'autre est le suffixe /-*k*/ du *maba*. Ici il y a une contradiction évidente.

Dans de nombreuses comparaisons, la plupart des exemples sont très courts, se composant d'un ou deux phonèmes. Voici, d'après Greenberg (1963:114), le pluriel commun hypothétique en /-*i*/ des langues chari-nilotiques : 1. nubien du Nil -*i* ; 2. *didinga* -*i* ; 6. *temein* -*i* ; *shilluk*, *acholi* (alur), *lango* -*i* ; *masai* -*i* ; *bari* -*e*. Hors des langues du Soudan oriental le suffixe -*i* se trouverait encore, selon lui, dans les langues suivantes : en *kunama* -*i* ; en *berta* -*e* ; en *moru-madi*, du groupe du Soudan central, comme -*e* et -*i* ; en *binga* -*i*, -*yi*, -*y*. Le *mangbetu* a aussi un -*i* pluriel dans les noms d'agent.

Néanmoins, ce qui est le moins persuasif, ce sont les thèses du système de classification de Greenberg, d'après lesquelles les groupes et les embranchements linguistiques sont rassemblés dans une plus grande famille d'une part, et les sous-groupes et les divisions ultérieures formées à l'intérieur de ces embranchements et familles d'autre part.

La connexité génétique des branches de la famille nilo-saharienne et des groupes de la branche chari-nilotique est étayée d'une argumentation de même caractère, qui figure aussi au cours de la démonstration de l'affinité de chaque langue de ces groupes. Dans ces passages le nombre des concordances hypothétiques ou réelles se multiplie de telle sorte qu'il cause un embarras de richesse sans pour autant renseigner sur la connexion qualitative des données et la relation hiérarchique de chaque groupe et sous-groupe de langues à l'intérieur des familles. Comment le lecteur peut-il croire que les groupes et sous-groupes sont dans la même corrélation génétique l'un avec l'autre comme le prétend Greenberg ?

De même, il faut souligner que les faits linguistiques contredisent ou paraissent contredire souvent les groupements de Greenberg. J'ai choisi deux rubriques (23. « grand » et 24. « souris ») de la comparaison nigéro-kordofanienne (1963:153-160) en groupant les faits de langues africaines dans deux types, conformément aux différences de la forme phonétique (naturellement c'est aussi un procédé très subjectif). J'ai marqué la lettre et le numéro du groupe de langues à côté de chaque forme (K = kordofanien, N = nigéro-congolais) :

	Type I	Type II	Type III	Type IV	Type V	Type VI
23. « Grand »	<i>faf</i> N1 <i>fifi</i> K4	<i>ba</i> N2 <i>bayi</i> N2 <i>boi</i> N1 <i>boli</i> N6 <i>bi</i> N3	<i>pei</i> N3 <i>pipri</i> N4	<i>opē</i> N4 <i>o:pi</i> K3 <i>opu</i> N6 <i>ipa</i> K1	<i>bo</i> N6	
24. « Souris »	<i>(ni)fa</i> K4 <i>(e)fia</i> N4 <i>(ni)fi</i> K4 <i>fife</i> K4 <i>foi</i> N1	<i>mfa</i> N5	<i>ofē</i> N4	<i>ipi</i> N5 <i>ipi</i> N5	<i>poi</i> K3 <i>powi</i> K3 <i>pou</i> K3	<i>pwe</i> N4

La variété des types, quand ils ne sont pas classés d'après les critères géographiques ou lexico-statistiques, mais d'après les données de la concordance phonétique, illustre bien la distance entre les langues du même groupe. Les mots de la rubrique 23, empruntée au diula et au limba, langues appartenant au même groupe linguistique nigéro-congolais et très proches parentes, ont soit un /f/, soit un /b/ initial, tandis que la forme du mot en tumtum (groupe kordofanien) concorde avec celle du diula (*faf-fifi*).

Dans les matériaux de Greenberg il y a beaucoup de données linguistiques qui sont discutables, obscures ou simplement erronées. Cela signifie que la documentation pourra modifier ses thèses aussi sous ce rapport, après avoir été contrôlée dans ses détails et épurée par les spécialistes des langues en question¹. Je vais citer quelques données controversables ou fausses.

Selon Greenberg (1963:14), l'équivalent de « mordre » est *dumu* en bambara. Prost (1953:81) cite ce mot sous la forme *du*, tandis que sa forme est *dé* dans la langue mwa. C'est probablement la donnée de Prost qui est exacte, d'autant plus que la justesse de cette forme est affirmée par d'autres faits des langues parentes. Dans ce cas la comparaison de Greenberg est douteuse parce que tous les autres mots cités ont une consonne nasale en deuxième position (mandjak *rume*, balante *doma*, etc.).

Chez Greenberg (1963:16), l'équivalent de « jour, soleil » est *la*, en dyula, tandis que le même mot est *de* en manon et en dan. Selon Prost (1953:112, 118), il y a deux mots différents qui désignent « période » et « corps céleste », à savoir, *nyene* et *nyènè* en mana, ou bien *dékpoy* et *nyě*, *nyâgě* et *nyan* dans les deux dialectes du mana. A vrai dire, ces formes seraient difficilement insérées parmi les mots comparés par Greenberg.

Selon le chercheur américain (1963:18), l'équivalent de « quatre » est *nani* en bambara. C'est le mot qui est admis dans son tableau. Par contre, Prost (1953:174) mentionne que ce mot n'est pas d'origine mandingue, c'est-à-dire qu'il est vraisemblablement emprunté. L'opinion de Prost est renforcée par le fait que nous trouvons des formes tout à fait différentes dans les dix-neuf langues examinées par cet auteur (à l'exception du bambara et du vaï, *nani*, et du bobo, *nā*) qui révèlent, par contre, une plus grande parenté entre elles : *sivo*, *si*, *siya*, *yiziè*, *ziè*, *singè*, *syén*, etc.

II. — DIFFÉRENCES ENTRE LA MÉTHODE COMPARATIVE ET CELLE DE LA « RESSEMBLANCE »

En dehors des questions de détail et des problèmes spéciaux concernant la classification des langues africaines, il faut examiner la justesse et la valeur de la théorie et de la méthode de Greenberg. Il applique essentiellement la méthode développée précédemment par Swadesh (1954) qui peut être nommée méthode de la « ressemblance », en face de la méthode comparative traditionnelle.

D'après le raisonnement de Greenberg, en comparant deux ou plusieurs langues dont la parenté linguistique n'est pas encore établie, on trouve toujours des mots identiques ou ressemblants sous le rapport de la forme phonétique et de la signification. Les motifs de cette concordance sont les suivants : il y a 1) des mots qui sont formés par symbolisme phonétique (onomatopée, expressivité, etc.), 2) des mots empruntés, 3) des mots d'origine commune et 4) des mots se ressemblant par pur hasard. Mais les cas d'emprunt sont très limités

1. Quelques autres critiques de Greenberg, comme par exemple TUCKER (1957 : 548-549), mentionnent aussi ces erreurs.

parce qu'ils se rencontrent surtout au milieu des mots caractéristiques de culture et de civilisation, et, si on fait abstraction de l'interprétation onomatopéique puisque de tels mots sont en général très rares et ne se rencontrent que dans une certaine sphère de notions, il ne reste pour explication que les cas de parenté ou de hasard. Mais si le nombre des mots ressemblants dépasse 8 %, le hasard peut être, à son tour, exclu de l'interprétation, c'est-à-dire que les langues examinées sont, dans ce cas, d'origine commune et issues de la même langue primitive. Greenberg n'explique pas de quoi découle cette limite de 8 %, mais il remarque que la proportion des cas de ressemblance aléatoire diminue rapidement, en proportion carrée, et devient pratiquement nulle si plus de deux langues sont comparées.

Avant de faire la critique de la méthode de ressemblance, je voudrais récapituler brièvement les principaux traits caractéristiques de la méthode comparative généralement reconnue et les possibilités d'application de cette méthode aux langues africaines.

La preuve de la parenté génétique des langues est la régularité des concordances de sons. Ces conformités régulières, appelées lois phonétiques, doivent se manifester, d'une part dans le vocabulaire et spécialement dans le vocabulaire fondamental, y compris les mots d'emprunt, d'autre part dans les éléments de la structure grammaticale. Les concordances régulières doivent être démontrées parmi les éléments lexicaux ou grammaticaux à signification ou fonction identiques ou ressemblantes, toute divergence de signification ou de fonction devant être vérifiée par des motifs de changements sémantiques ou fonctionnels. Par contre, la ressemblance ou même la correspondance régulière des pronoms ont une force probante minime, et l'origine commune des pronoms reste douteuse même si les autres éléments des langues en question soutiennent l'épreuve de la comparaison. La cause de ce phénomène est, d'une part que les pronoms sont en général très courts, et que beaucoup d'entre eux prennent leur origine du symbolisme phonétique d'autre part. Greenberg, comme Delafosse (Meillet et Cohen, 1952), se réfère souvent à la correspondance des pronoms sans tenir compte du fait que les formes phonétiques de bien des pronoms coïncident presque parfaitement dans la plupart des langues du monde, qu'elles soient parentes ou non.

Il s'ensuit des critères précédents que la structure phonétique et grammaticale des langues comparées ne doit pas forcément être analogue, ni les mots confrontés identiques ou ressemblants, et que, en général, la parenté génétique ne signifie nullement que les langues apparentées se ressemblent, mais seulement qu'elles ont une origine commune et que les deux états peuvent être vérifiés par des concordances régulières, c'est-à-dire par des lois phonétiques. Il y a des différences énormes entre les structures phonétiques, grammaticales et lexicales du français, de l'allemand, du grec moderne, du latin, du hindi : ce sont pourtant des langues parentes. Et comme Meillet (1954:32, 1^{re} éd. en 1925) le remarque dans son exemple fameux, le mot arménien *evku* « deux » a la même origine que les autres mots de même signification des langues indo-européennes (*duo*, *dva*, *two*, etc.), quoique aucun de ses phonèmes ne « ressemble » aux leurs.

Je voudrais rappeler encore quatre difficultés que l'on rencontre habituellement en appliquant la méthode comparative.

La première difficulté se pose quand il y a des exceptions — quelquefois beaucoup d'exceptions — aux concordances régulières : l'existence d'au moins une loi phonétique incontestable, qui se manifeste aussi bien dans le lexique que dans la structure grammaticale, reste tout de même un critère indispensable de la parenté génétique de deux ou plusieurs langues.

La deuxième difficulté se présente quand la série de comparaisons comporte une forte proportion d'éléments trop courts, comprenant seulement de un à trois

phonèmes : dans le cas, par exemple, du suffixe *-i* du pluriel des langues indo-européenne et finno-ougrienne, ainsi que du morphème identique de certaines langues négro-africaines mentionné par Greenberg (1963:114).

La troisième difficulté se rencontre dans le cas où il y a peu de morphèmes grammaticaux dans les langues comparées, comme par exemple dans les langues soudano-guinéennes isolantes.

Enfin, la quatrième difficulté se manifeste dans le cas où les langues apparentées diffèrent très peu l'une de l'autre, puisque alors ni le groupement exact ni la relation précise de ces langues ne peuvent être vérifiés par des moyens linguistiques. Cela concerne par exemple plusieurs langues ou dialectes bantu.

Quant à la réalité de la limite de 8 % fixée par Greenberg, j'ai essayé de la contrôler par une méthode exacte. Au cours de ce travail, j'ai fait opérer des calculs dans trois langues artificielles de type africain sur la base des données du peul, du mwa et du songhai¹. D'après les calculs, le nombre des mots ressemblants ou identiques par hasard s'échelonne de 7,99 à 10,02 % pour des langues possédant un vocabulaire fondamental de 5 000 mots.

Je contrôlerai maintenant les calculs de Greenberg basés sur la sommation mécanique des comparaisons d'unités de mots isolées, par une méthode qui tient compte d'un vocabulaire fondamental de 5 000 mots pour chaque langue.

J'ai divisé le vocabulaire fondamental en quatre ou cinq couches. La couche 1 se compose des mots dissemblables, la couche 2 comprend les mots pareils par hasard, la couche 3 renferme les mots à symbolisme phonétique (onomatopée, idéophone, expressivité, etc.)², la couche 4 se compose des mots empruntés à d'autres langues africaines³. La couche 5 apparaît dans le cas où les langues comparées sont apparentées. Les mots d'origine commune et ressemblants appartiennent alors à cette couche, tandis que les mots d'origine commune mais présentant des formes divergentes dues aux changements phonétiques doivent être renvoyés dans la couche 1.

Dans mes calculs, j'ai fixé la proportion de la couche 3 à 15 %, celle de la couche 4 à 45 % par approximation⁴. En comparant des langues parentes, nous fixons la couche 5, également par approximation mais de deux manières différentes. Dans le cas où la séparation des deux langues parentes est soit de date ancienne, soit de date récente mais avec des changements linguistiques très considérables, la proportion de la couche 5, c'est-à-dire celle des mots d'origine commune et ressemblants est fixée à 3 %, tandis que dans le cas où, au contraire, la séparation est soit relativement récente, soit ancienne mais avec très peu de changements linguistiques, la proportion est fixée à 20 %⁵.

Comme les chiffres obtenus par les calculs mathématiques sont à peu près conformes à la valeur de 8 % déterminée par Greenberg par une méthode qui m'est inconnue, j'ai fixé arbitrairement la mesure de la couche 2, c'est-à-dire la proportion des mots se ressemblant par hasard, à 8 %.

1. Les calculs ont été faits par le mathématicien György Bánkóvi, chercheur scientifique de l'Institut de Mathématiques de l'Académie des Sciences de Hongrie.

2. En opposition avec l'opinion de Greenberg, il est manifeste que cette couche ne peut pas être négligée. C'est un fait évident pour tous les africanistes et qui n'exige pas une explication plus détaillée dans cet article.

3. Cette couche est sans doute très importante dans les langues africaines, de même que dans d'autres langues qui sont en relation étroite avec les langues voisines, comme par exemple les langues balkaniques.

4. Dans la publication anglaise de mon étude j'ai détaillé les motifs sur la base desquels ces chiffres ont été fixés.

5. Dans mon étude originale, je me réfère à quelques analogies avec d'autres familles de langues.

Voici maintenant les modèles des trois cas théoriques de la comparaison génétique des langues africaines :

1. *Relation sans parenté génétique.*

Couche 1	32 %	1 600 mots
Couche 2	8 %	400 mots
Couche 3	15 %	750 mots
Couche 4	45 %	2 250 mots
TOTAL	100 %	5 000 mots

Le nombre des mots dissemblables est de 1 600, c'est-à-dire 32 %, et celui des mots ressemblants (par hasard, par emprunt et par symbolisme phonétique) est au total de 3 400 ou 68 %.

2. *Parenté génétique ancienne ou séparation plus récente de langues très transformées.*

Couche 1	29 %	1 450 mots
Couche 2	8 %	400 mots
Couche 3	15 %	750 mots
Couche 4	45 %	2 250 mots
Couche 5	3 %	150 mots
TOTAL	100 %	5 000 mots

Le nombre des mots dissemblables est de 1 450 ou 29 %, tandis que celui des mots ressemblants (par hasard, par emprunt, par onomatopée et mots d'origine commune gardant la ressemblance) est de 3 550, c'est-à-dire 71 %.

3. *Séparation de langues parentes à une date récente ou séparation ancienne de langues archaïques.*

Couche 1	12 %	600 mots
Couche 2	8 %	400 mots
Couche 3	15 %	750 mots
Couche 4	45 %	2 250 mots
Couche 5	20 %	1 000 mots
TOTAL	100 %	5 000 mots

Les mots dissemblables sont 600 ou 12 % et les mots ressemblants sont 4 400, c'est-à-dire 88 %.

Il résulte de ces modèles théoriques basés sur certaines analogies que la méthode de ressemblance est absolument inutilisable parce qu'elle fait défaut justement là où on en aurait besoin. Cette méthode est incapable de distinguer

entre les langues caractérisées par des relations de type 1 et 2, car la différence entre elles n'est plus que de 150 mots, c'est-à-dire de 3 %. Certes, la comparaison des langues du type 3 pourrait être fructueuse, mais ce cas n'est pas important dans la pratique : il concerne des langues qui non seulement sont visiblement parentes, même aux yeux des profanes, mais qui sont plus ou moins compréhensibles entre elles, comme les langues néo-latines, les langues slaves, quelques groupes bantou et les langues tagoi, tumèli et talodi.

On pourrait objecter à mon argumentation qu'il y a des transitions graduelles entre les modèles 2 et 3, où la proportion de la couche 5 s'approche plus du type 1, ou bien où le type 2 s'éloigne plus du type 1. Naturellement de tels cas peuvent exister, mais après, selon les calculs, seule la proportion des mots ressemblants par hasard (8 %) pourra être tenue pour certaine ; en dehors de cela on peut déterminer, sur la base de critères très subjectifs, la proportion des mots dissemblables, mais rien n'est connu concernant les autres couches : par conséquent, la proportion des mots d'emprunt peut être de 60 % ou plus (c'est le cas de l'albanais parmi les langues indo-européennes) ; il est absolument impossible d'établir si la couche 5 existe ou non, donc si la parenté génétique peut être admise ou non, et, dans le cas affirmatif, quelle en est la mesure chiffrée et quels sont les mots qui y appartiennent.

Nous venons de dire que la quantité de la couche 2, soit 8 %, est connue. Mais il faut faire une restriction considérable à cet égard. Il est vrai que le facteur d'incertitude dû à la différence de fréquence des phonèmes dans les langues examinées peut être éliminé par des calculs statistiques, mais Greenberg n'a pas fait de calcul de cette sorte, c'est pourquoi des déviations de 2 à 3 % doivent être prises en considération dans ses comparaisons. Mais il y a une incertitude plus importante, à savoir la fluctuation des coefficients de probabilité. Bien que je n'aie pas fait calculer les déviations statistiques vu la longueur de ce procédé, on peut — sur la base d'approximations — fixer la limite supérieure de ces valeurs à 13-15 % environ ; par conséquent la déviation moyenne peut être plus grande que la différence des couches hypothétiques en question du type 1 et 2.

Dans les calculs, deux langues seulement ont été comparées en même temps, et les pourcentages des mots identiques ou ressemblants par hasard ont été obtenus conformément à ce procédé. Il est évident — et Greenberg a raison en l'affirmant — que dans le cas de comparaison entre plus de deux langues en même temps, ce pourcentage diminue de telle façon qu'on peut le négliger en fin de compte. D'après Greenberg, les ressemblances qui subsistent ensuite ne peuvent être expliquées autrement que par la parenté génétique.

Nous avons cependant deux objections contre les arguments de Greenberg. Premièrement, suivant notre modèle, le cas en question ne peut être évalué autrement qu'en réduisant la valeur de 8 % de la couche de telle sorte qu'elle puisse atteindre pratiquement le zéro. Comment ce cas se produit-il ? Sans doute, le pourcentage de la couche 1 augmentera jusqu'à 8 % inclusivement et alors la situation sera la suivante :

TYPE I.

Couche 1	40 %	2 000 mots
Couche 2	0 %	0 mot
Couches 3-4	60 %	3 000 mots
TOTAL	100 %	5 000 mots

TYPE 2.

Couche 1	37 %	1 850 mots
Couche 2	0 %	0 mot
Couches 3-5	63 %	3 150 mots
TOTAL	100 %	5 000 mots

TYPE 3.

Couche 1	20 %	1 000 mots
Couche 2	0 %	0 mot
Couches 3-5	80 %	4 000 mots
TOTAL	100 %	5 000 mots

C'est-à-dire qu'il est impossible de séparer les couches 3-5 par la suite, et la différence entre les trois types s'estompe encore davantage par rapport au cas de la comparaison de deux langues seulement.

Deuxièmement, dans le cas de la comparaison de plusieurs langues — si elles dérivent directement de la même langue primitive —, la proportion de l'héritage lexical commun diminue également parce qu'une faible proportion seulement des mots qui existaient dans la langue primitive subsiste dans toutes les langues dérivées. J'allègue l'exemple des données de Bereczki (1963) concernant les langues finno-ougriennes. L'auteur a examiné le vocabulaire fondamental du mari (tchérimisse) pour savoir combien de radicaux finno-ougriens de l'héritage commun ont été conservés par comparaison avec les autres langues finno-ougriennes. Selon les données de Bereczki, il y a 694 radicaux finno-ougriens en mari, donnant :

283	mots indubitablement communs avec le mordvine
346	— — — — ud-murt (votiak)
255	— — — — handi (ostiak)
240	— — — — vogoul
196	— — — — hongrois

Ces données indiquent non seulement le fait que l'héritage finno-ougrien atteint à peine 10 % du vocabulaire du mari, qui comprend plusieurs milliers de radicaux, mais aussi que les mots communs avec le mordvine ne s'élèvent qu'au tiers de cet héritage.

Il s'ensuit de tout cela que, dans le cas de comparaison de plusieurs langues, non seulement la couche 2 diminuera, disparaîtra même, mais, inévitablement, la couche 5 aussi, surtout s'il s'agit d'une parenté ancienne ou bien de langues très changeantes, tandis que, par contre, la valeur de la couche 1 augmentera. La situation est à peu près la suivante :

TYPE 1.

Ne change pas, c'est-à-dire la proportion des deux sortes de couches est de 40 % à 60 %, ou 2 000 mots à 3 000 mots.

TYPE 2.

Couche 1	39 %	1 950 mots
Couche 2	0 %	0 mot
Couches 3-5	61 %	3 050 mots
TOTAL	100 %	5 000 mots

TYPE 3.

Couche 1	25 %	1 250 mots
Couche 2	0 %	0 mot
Couches 3-5	75 %	3 750 mots
TOTAL	100 %	5 000 mots

On peut constater que l'observation de Greenberg n'a point amélioré la valeur de la méthode de ressemblance car la différence par couches entre les trois types s'estompe encore davantage.

Certes, les langues parentes peuvent non seulement dériver de la même langue primitive mais présenter en outre une affinité plus indirecte et plus compliquée, comme par exemple la parenté de certaines langues finno-ougriennes (hongrois, vogoul, ostiak) qui sont dérivées de la branche ougrienne, c'est-à-dire qui sont plus proches les unes des autres qu'elles ne le sont de la langue mari (tchéremisse). Il est donc peu vraisemblable que ces langues gardent autant d'éléments communs que dans le cas d'une parenté plus proche. Donc, si le nombre des éléments communs est moindre, cela ne signifie pas infailliblement que nulle parenté n'existe (en supposant que le nombre des éléments communs signifie quelque chose, ce qui n'est rien moins que prouvé), mais que la parenté est plus éloignée. Dans la comparaison des langues d'Afrique noire, il est aussi peu probable qu'un groupe de dix à quinze langues remonte à la même langue primitive, même si elles sont parentes au fond. Il est plus vraisemblable que quelques-unes de ces langues sont parentes plus proches et composent un groupe séparé se rattachant, par concaténation, aux autres langues. Mais, si les langues examinées ont une relation de ce type, la couche des mots véritablement communs doit être mince, ne comptant pas beaucoup plus de quelques centaines de mots. La méthode de ressemblance est donc tout à fait inadéquate pour démontrer et séparer numériquement cette couche des autres, tout comme elle est insuffisante à détacher les couches 3 et 4 l'une de l'autre.

CONCLUSIONS

Ci-dessus j'ai déjà tiré la conclusion méthodologique principale que je voudrais répéter ici : sans la pierre de touche des lois phonétiques les rapports d'affinité des langues négro-africaines ne peuvent être prouvés et la classification génétique ne peut en être effectuée sans l'application de ces principes. Autrement dit, même des théories aussi séduisantes que celle de Greenberg ne dépassent point le niveau des hypothèses de travail qui seront relayées par d'autres hypothèses basées sur une documentation plus solide. La théorie de Greenberg a écarté les hypothèses précédentes, parce que plus méthodique et approfondie. Mais la sentence de Gabelentz (1901:158) s'applique également à toutes ces tentatives : « *Sprachvergleichung ohne Lautvergleichung ist gedankenlose Spielerei.* »

Le deuxième enseignement est aussi indiscutable : pour le moment, nulle synthèse tendant à l'exhaustivité, comme celle de Greenberg, ne peut être considérée avec sérieux tant que nous ne posséderons qu'une connaissance tellement incomplète et inégale de la Babel africaine. Bien heureux encore, si dans les glossaires, un équivalent de la langue africaine se trouve en français, en anglais ou dans une autre langue européenne ; mais les synonymes, le champ lexical entier restent inconnus dans la plupart des langues africaines. Pour apprendre le chemin des mots d'emprunt, des œuvres lexicographiques solides et des recherches modernes de géographie linguistique seraient indispensables. C'est pourquoi — avant la compilation de synthèses prétentieuses — les lacunes de notre connaissance devraient être comblées par les investigations des spécialistes de chaque langue et groupe de langues négro-africaines¹. Au demeurant, il n'est pas certain que, même en possession de toutes les données synchroniques, la connaissance de l'histoire des langues africaines serait assez complète pour une classification génétique digne de confiance, faute de monuments linguistiques suffisamment nombreux et étendus. La méthode historique proprement dite étant inapplicable, seule la reconstruction interne viendra à l'aide de la méthode comparative. Avant qu'on en soit là, les africanistes ont encore beaucoup de tâches à accomplir.

Budapest, juin 1967.

BIBLIOGRAPHIE

BERECZKI, G.

- 1963 « *Vzaimootnošenie marijskoj leksiki s leksikoj mordovskih i permskih jazыkov* » [Rapport réciproque entre le lexique du tchérémisse et celui du mordvine et des langues permiennes], *Congressus Internationalis Fenno-Ugristarum Budapestini Habitus*, 20-24 septembre 1960 [redigit G. Ortutsay], Budapest.

GABELENTZ, G. VON DER

- 1901 *Die Sprachwissenschaft*, Leipzig.

1. Il faut mentionner la monographie solide et moderne de THOMAS (1963) comme un exemple des recherches nécessaires pour la connaissance profonde des langues de l'Afrique noire.

GREENBERG, J. H.

- 1957 "Nilotic, 'Nilo-Hamitic', and Hamito-Semitic: A Reply", *Africa*, XXVII, 4, pp. 364-378.
1963 *The Languages of Africa*, La Haye.

MEILLET, A.

- 1954 *La méthode comparative en linguistique historique*, Oslo (1^{re} éd., 1925).

MEILLET, A. et COHEN, M.

- 1952 *Les langues du monde*, Paris.

PROST, A.

- 1953 *Les langues mandé-sud du groupe mana-busa*, MIFAN, 26, Dakar.

SWADESH, M.

- 1954 "Perspectives and Problems of Amerindian Comparative Linguistics", *Word*, 10, pp. 306-332.

THOMAS, J. M. C.

- 1963 *Le parler ngbaka de Bokanga : Phonologie, morphologie, syntaxe*, Paris - La Haye.

TUCKER, A. N.

- 1957 "Philology and Africa", *Bulletin of the School of Oriental Studies*, 20, pp. 541-554.

VOEGELIN, C. F. and F. M.

- 1964 "Languages of the World", *Africa* 1, *Anthropological Linguistics*, VI, 5, Bloomington.