

## UN ESSAI DE CLASSIFICATION LEXICO-STATISTIQUE DES LANGUES DE LA FAMILLE TCHADIQUE PARLEES AU CAMEROUN

Daniel BARRETEAU

(ORSTOM - MESRES)

Dans le cadre du programme d'Atlas Linguistique du Cameroun (ALCAM), sur un corpus assez conséquent (57 langues) et homogène (des listes de 120 mots relevées en grande partie par le même chercheur), nous avons tenté une première exploitation lexico-statistique à base de jugements de ressemblance, l'originalité de la méthode résidant dans la reconnaissance de plusieurs degrés de ressemblance, ce qui a permis, par la suite, de procéder à plusieurs modes de calcul pour les regroupements en arbres classificatoires.

### 1. JUGEMENTS DE RESSEMBLANCE

1.1. Notre méthode consiste à appliquer des taux de ressemblance différenciés ou "pondérés" selon que nous jugeons que les ressemblances (ou les différences) sont plus ou moins prononcées entre les formes comparées. En d'autres termes, pour calculer la proximité de deux langues, nous attribuons plus de points de ressemblance lorsque deux formes comparées sont identiques ou très proches l'une de l'autre que lorsqu'elles ne se ressemblent que sur de rares points.

Nous choisissons, dans une échelle à six degrés (5, 4, 3, 2, 1 / 0), l'un ou l'autre coefficient selon le degré de ressemblance, 5 étant le maximum et 0 le minimum.

Nous évitons ainsi les jugements binaires du type :

1. les formes se ressemblent et doivent être apparentées /
0. les formes sont différentes et ne peuvent être apparentées, jugements très difficiles à opérer de façon régulière et très aléatoires à défaut d'une étude comparative approfondie. Avec cette méthode, les différences d'appréciation entre linguistes peuvent être considérables surtout lorsque les langues sont très éloignées et qu'on ne bénéficie d'aucune connaissance supplémentaire sur le corpus en question.

Selon notre méthode, nous pouvons être plus ou moins "sévères", hésiter entre tel et tel autre degré de ressemblance, mais non pas sauter de deux, trois ou quatre degrés, ce qui garantit une certaine cohérence.

Après plusieurs essais sur les langues qui nous intéressent, nous avons défini cette grille de la façon suivante :

- Coefficient 5 : identité totale, aucune différence ou seulement une différence phonétique minime comme par exemple une différence de degré d'aperture dans les voyelles (opposition non-pertinente dans certaines langues et certains contextes) : ainsi, une différence entre i/e, ə/a ou u/o, est généralement considérée comme nulle à ce niveau. Des différences vocaliques sont considérées comme négligeables dans la catégorie du verbe où le statut des voyelles est difficile à cerner.

Les différences conditionnées par le contexte se situent également dans ce degré : ainsi, les différences entre les consonnes alvéolaires (ts, dz, s, z) et les consonnes palato-alvéolaires (c, j, sh, zh) sont considérées comme nulles si elles apparaissent conditionnées par l'environnement vocalique ; des séquences avec consonne postérieure labialisée suivie d'une voyelle non-arrondie sont équivalentes à des séquences avec postérieure non-labialisée suivie

de voyelle arrondie : kwa = ko , kwe = kœ .

- Coefficient 4 : une différence phonétique entre des phonèmes assez proches, c'est-à-dire des phonèmes qui présentent un ou deux traits de différence comme par exemple une différence dans le mode d'articulation (occlusive / affriquée / fricative, occlusive dentale / latérale / vibrante, nasale / orale) ou une différence quant au trait de voisement (sourde / sonore).

Dans cette catégorie nous rangeons également des différences d'étoffement simple des radicaux : segment initial (voyelle initiale / voyelle  $\emptyset$ ) ou final (consonne finale "faible", considérée comme un suffixe / consonne  $\emptyset$ ), ou des différences de prosodie (palatalisation, labialisation).

- Coefficient 3 : deux différences phonétiques cumulées : soit un phonème du radical nettement différencié, soit deux phonèmes légèrement différents ; ou bien une différence morphologique évidente (suffixe ou préfixe différent) sans modification interne des formes comparées (même composition et même ordre des consonnes radicales).

Dans ce degré nous situons également les cas de métathèse facilement perceptibles : inversion de deux éléments sans autre modification majeure.

- Coefficient 2 : différenciation accentuée des items bien qu'ils restent comparables. Cela peut se traduire par des changements simples sur plusieurs phonèmes. Cela peut se traduire aussi par une différence morphologique plus prononcée qu'au niveau 3, par exemple par une métathèse avec, en plus, d'autres changements phonétiques.

- Coefficient 1 : les différences prévalent sur les points de ressemblance. La ressemblance est peu perceptible au vu des deux seules formes en présence mais, compte-tenu de certaines règles de correspondance, de la "chaîne des réflexes" (principe de transitivité), on suppose cependant que les deux mots peuvent remonter à une même racine.

- Coefficient 0 : différence lexicale, aucun point de

ressemblance entre les deux formes en présence, même en tenant compte des réflexes dans les langues voisines. Les deux items ne paraissent pas issues d'une même racine.

1.2. L'application de cette grille ne va pas sans poser de problèmes, surtout dans les groupes constitués d'un grand nombre de langues, comme le groupe mafa, qui comporte 19 langues, où il faut essayer de maintenir une certaine homogénéité dans les jugements. L'opération ne peut se faire en une seule journée !

Que l'on songe que nous atteignons un total de 26.600 jugements de ressemblance, pour les comparaisons internes aux groupes :

groupes	nombre de langues	nombre de jugements
masa	3	3 x 100 : 300
wandala	6	15 x 100 : 1 500
mafa	19	171 x 100 : 17 100
higi	10	45 x 100 : 4 500
daba	5	10 x 100 : 1 000
mida'a	2	1 x 100 : 100
mandage	7	21 x 100 : 2 100
total	52	266 x 100 : 26 600

Etant donnée cette masse considérable de jugements et la complexité des données à comparer, il est bien évident qu'il est extrêmement difficile, voire impossible, d'appliquer d'une façon invariable et "sans faute" la grille théorique proposée. Cependant, pour essayer d'obtenir la plus grande homogénéité possible, il peut être envisagé de faire plusieurs contrôles, rendus possibles à partir du moment où nous publions l'intégralité des listes comparatives et où nous conservons noir sur blanc chaque jugement de ressemblance (au lieu de publier seulement les résultats comme cela se fait généralement). Les différentes démarches sont les suivantes :

a) On analyse les formes attestées pour un même item ; on tente d'établir des regroupements et des pseudo-reconstructions en s'inspirant éventuellement de reconstructions

déjà proposées par d'autres auteurs. Chaque item est classé dans un type. Cette démarche donne une première vue d'ensemble, que l'on pourra modifier par la suite.

b) On effectue les jugements de ressemblance en passant en revue toutes les langues d'un même groupe, item après item ; soit, par exemple dans le groupe masa, qui comporte trois langues (sans compter le zumaya), l'item n°1 "bouche" : on compare la forme du masa à celles du musey et du lame, puis celle du musey à celle du lame ; on examine ensuite l'item n°2 "oeil" pour les mêmes langues. A la fin de la liste, on passera à un autre groupe de langues. Cette démarche permet de regrouper les formes semblables à l'intérieur de sous-groupes et d'étalonner les jugements.

c) On effectue les jugements de ressemblance d'une langue par rapport aux autres langues de son groupe en passant en revue directement les cent mots de cette langue avant de passer à la langue suivante. Cette démarche permet de découvrir certains aspects de la phonologie et de la morphologie d'une langue tout en révélant parfois des règles de correspondance avec les autres langues du groupe.

d) On examine la liste complète des paires de mots entrant dans un même degré de ressemblance.

e) On invite un autre linguiste à appliquer la même méthode sur le même corpus.

f) On traite les mêmes données avec des moyens informatisés.

g) On compare les résultats globalement et dans le détail. Soulignons ici qu'il serait absurde de comparer des résultats portant sur quelques mots seulement et sur des langues en nombre très limité. Les résultats obtenus ne valent pas tellement en absolu mais plutôt en fonction d'un ensemble.

1.3. Quelques principes secondaires viennent nuancer l'application de notre grille :

a) Notre méthode se situe sur un plan synchronique et ne nécessite pas nécessairement une comparaison développée de tout le corpus avec des reconstructions historiques - ce qui serait difficile sur un corpus d'une centaine de mots -. Cela entraîne que l'on peut accorder des coefficients nuls à des paires de mots qui remonteraient pourtant à un même étymon ou, inversement, des coefficients positifs à des mots non-apparentés. Le risque est certain. Cependant, nous pouvons imaginer, avec la loi des grands nombres, que les erreurs dans l'un ou l'autre sens se corrigent plus ou moins mutuellement et donc que les résultats de nos jugements "rapides", immédiats, peuvent apporter des indications utiles sur l'état actuel des langues en présence, sur le plan lexical. L'hypothèse est que cet état doit refléter plus ou moins la filiation génétique de ces langues, cette méthode servant de première approche pour une étude comparative historique. Nous reconnaissons, bien sûr, que seule la méthode de reconstruction historique, avec l'établissement de règles de correspondances, permet de décider de l'apparentement génétique ou non de deux langues.

b) Le fait que les jugements se font toujours sur un couple de formes entraîne que l'on ne devrait pas tenir compte du principe de transitivité :

si A ressemble à B  
et si B ressemble à C  
alors A ressemble à C.

Selon notre méthode, nous considérons cependant le principe de transitivité pour accorder un coefficient positif minimum (1) plutôt que d'attribuer un coefficient nul (0) ; ceci est particulièrement important pour les langues très éloignées où les ressemblances sont rarement manifestes, où le recours à des formes tierces ou formes reconstruites est nécessaire pour établir la cognation.

c) Lorsque des règles de correspondance apparaissent,

nous les signalons et en tenons compte dans les jugements par un moindre degré de "sévérité" : par exemple, là où nous devrions porter un coefficient 3, si une règle de correspondance est établie, nous attribuons plutôt un coefficient 2.

d) Comme généralement dans les comparaisons de ce type, on accorde plus de poids aux consonnes (ou "squelette consonantique") qu'aux voyelles. Ceci est d'autant plus valable pour les langues de la famille tchadique où la charge lexicale se porte davantage sur les consonnes et les faits grammaticaux sur les voyelles.

e) Il est à noter que certains mots, particuliers à cette famille de langues, se prêtent beaucoup plus que d'autres à notre méthode de calcul ; ainsi les formes pour les items "bouche", "oreille", "lune", "soleil", "viande", "éléphant", "poisson", "deux", "trois", "mourir", "savoir", "boire", etc., sont aisément comparables : il est possible de compter le nombre de traits différents ce qui fait que nos jugements de ressemblance sont ici très précis et peuvent être plus "sévères" étant donné que les formes sont moins différenciées. En revanche, d'autres mots comme "oeil", "oeuf", "tortue", "enfant", etc., présentent des écarts formels considérables (en particulier avec des redoublements, préfixations, suffixations, métathèses...) si bien que dans ces cas-là, les jugements sont évidemment plus difficiles à effectuer et sont, de fait, moins "sévères".

f) Les segments internes ont plus d'importance que les segments périphériques, tout le problème étant de distinguer les éléments de base des affixes et de catégoriser les développements morphologiques spécifiques pour chaque langue ou groupe de langues, ce qui ne va pas de soi avec un corpus aussi limité.

g) Ce que l'on considère comme des différences nulles dans cet essai sur 57 langues pourrait être considéré comme des différences significatives dans une étude dialectométrique plus fine et plus complète, portant sur des corpus

plus étendus (500 ou 1 000 mots).

h) Dans une paire de mots à comparer, il faut tenir compte du rapport entre les éléments communs et les éléments différents. Nous situerons, par exemple, la paire kwalakwam / kwalakwad', où le rapport de ressemblance est de 6 phonèmes sur 7, dans le degré 4, bien que les consonnes /m/ et /d/ soient très éloignées. En revanche, on sera beaucoup plus sévère pour de simples différences sur des formes courtes où l'on aurait, par exemple, des rapports de ressemblance de 1 sur 2 ou de 1 sur 3.

i) Lorsqu'une entrée se traduit par deux formes (voir, par exemple, "enfant" au singulier et au pluriel), on conserve les deux formes et on choisit celle qui est la plus proche de celle à comparer dans l'autre langue. Il faut souligner que cette situation est très peu fréquente.

Ce biais à la méthode stricte de la lexico-statistique permet, en quelque sorte, de compenser les cas "d'erreurs" dans les traductions, erreurs inévitables lorsqu'on dispose d'aussi peu d'éléments sur des langues totalement inconnues par ailleurs, erreurs qui ne font que baisser, anormalement, les taux de ressemblance.

j) Nous comparons des formes structurelles, débarrassées des affixes manifestes comme par exemple les suffixes de genre en masa, musey, etc., les préfixes ou suffixes d'infinitif, etc. Ces marques seraient à considérer dans une étude comparative portant sur la grammaire.

Dans certaines langues mal connues, il est parfois bien difficile de séparer les affixes des radicaux.

1.4. Nous avons pu répartir les 57 langues étudiées en 11 groupes, au simple vu des dix ou vingt premiers mots.

Cette première répartition sera réexaminée par la suite. Nous choisirons une ou plusieurs langue(s) représentative(s) de chaque groupe et sous-groupe pour calculer de nouveau les distances avec les autres langues représentatives des autres groupes.



Parmi les 57 langues, si l'on choisit 23 langues représentatives des groupes et sous-groupes, nous aurons encore à effectuer :

191 x 100 : 19 100 jugements de ressemblance  
pour définir les rapports entre les groupes. Soit au total :  
26 000 + 19 100 : 45 700 jugements de ressemblance.

Si nous devions calculer l'intégralité des rapports entre les 57 langues, nous atteindrions la somme de 123 100 jugements de ressemblance ! Cela ne serait ni utile, ni possible manuellement.

Au stade actuel de nos recherches, nous n'avons pas encore calculé les distances entre les différents groupes de langues.

De même, la position de certaines langues n'est pas encore bien établie : les enquêtes sur le baldamu et le gaduwa (groupe mafa) sont récentes ; nous ne disposons que d'une dizaine de mots sur le zumaya (groupe masa), langue quasiment disparue ; il se peut que quelques langues éparses (ou dialectes) n'aient pas été relevé(e)s dans le groupe mida'a (région de Zina, Ma'é).

1.5. Sur la liste de 120 mots employée pour l'Atlas Linguistique du Cameroun, nous avons écarté 20 mots peu satisfaisants pour la comparaison. Dans les traductions nous y avons relevé des réflexes multiples, même entre langues très proches, des composés, des emprunts.

Les mots écartés sont les suivants : fesse, aile, nuage, rivière, fumée, serpent, pou, écorce, feuille, sel, six, sept, huit, neuf, voler (s'envoler), dire.

D'autres mots ont été regroupés : plume - poil, personne - homme, marcher - partir, frapper - tuer.

## 2. CALCULS

2.1. Lorsque tous les jugements de ressemblance sont effectués, nous totalisons le nombre de paires (ou de jugements) entrant dans chaque coefficient. Ainsi, par exemple, entre les langues masa et musey, nous avons dénombré :

degré 5	:	45 paires d'items
degré 4	:	22 paires d'items
degré 3	:	5 paires d'items
degré 2	:	9 paires d'items
degré 1	:	2 paires d'items
degré 0	:	17 paires d'items

2.2. Ensuite, nous regroupons les degrés 5 et 4 pour obtenir un "taux minimal de ressemblance" ou "taux de ressemblance manifeste".

Taux minimal de ressemblance entre masa et musey :  
45% + 22% : 67%

En regroupant les degrés 5, 4, 3, 2 et 1, nous obtenons un "taux maximal de ressemblance" ou "taux de ressemblance profonde" :

45% + 22% + 5% + 9% + 2% : 83 %

C'est ce taux de ressemblance qui définit le nombre de racines supposées communes et qui devrait être employé dans des calculs glottochronologiques.

Un "taux pondéré de ressemblance" sera obtenu en appliquant les divers coefficients et en réduisant en pourcentage :

$$\frac{(45 \times 5) + (22 \times 4) + (5 \times 3) + (9 \times 2) + (2 \times 1)}{5} = 69 \%$$

Ce dernier taux répond précisément à la méthode d'affinement des jugements de ressemblance que nous préconisons.

2.3. Pour chaque groupe de langues, nous pouvons ainsi fournir trois matrices de ressemblance lexical avec taux minimal,

taux pondéré, taux maximal.

Pour le groupe masa, nous obtenons les résultats suivants :

Taux minimal	masa	musey	lame
masa	-	67	34
musey	67	-	35
lame	34	35	-

Taux pondéré	masa	musey	lame
masa	-	69,8	44,4
musey	69,8	-	46,6
lame	44,4	46,6	-

Taux maximal	masa	musey	lame
masa	-	83	65
musey	83	-	69
lame	65	69	-

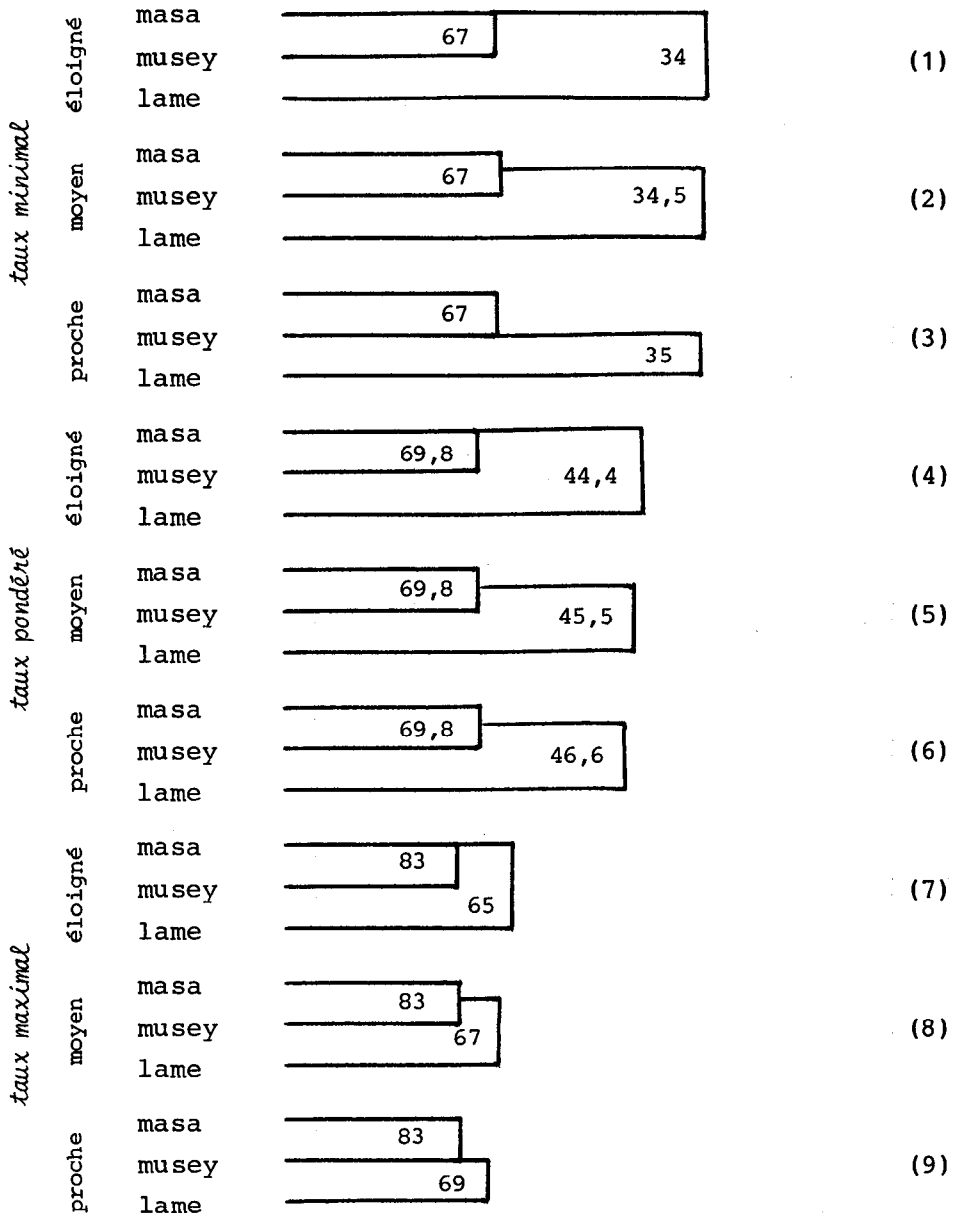
2.4. Pour la technique des regroupements sous forme d'arbres, nous employons également trois méthodes :

- le voisin le plus éloigné : pour réunir deux groupes de langues, par exemple A-B et C-D (A et B, C et D sont déjà réunies car elles sont mutuellement les plus proches), on sélectionne les taux de ressemblance les plus faibles, par exemple B-D, si  $B-D < B-C < A-D < A-C$ .

- la distance moyenne est la moyenne arithmétique des distances :  $(A-C) + (A-D) + (B-C) + (B-D)$

- le voisin le plus proche : on sélectionne le taux de ressemblance le plus fort, par exemple A-C, si  $A-C > A-D > B-C > B-D$ .

2.5. Ainsi, pour chaque groupe de langues nous obtenons neuf arbres classificatoires comme on peut le montrer avec le groupe masa :



2.6. L'expérience montre que les neuf arbres ont souvent la même configuration, avec une diminution proportionnelle des écarts :

arbres	masa / musey	masa-musey / lame
(1)	67	34
(2)	67	34,5
(3)	67	35
(4)	69,8	44,4
(5)	69,8	45,5
(6)	69,8	46,6
(7)	83	65
(8)	83	67
(9)	83	69

L'intérêt de la procédure réside dans les cas de configurations divergentes des différents arbres d'un même groupe, comme on pourra le constater dans le groupe wandala, en annexe, où les positions du parakwa et du gavoko ne sont pas évidentes.

2.7. Pour départager les cas problématiques et proposer, *in fine*, une classification unique et simplifiée, nous avons retenu les taux pondérés de ressemblance que nous avons ramenés à la dizaine la plus proche (la demi-dizaine serait peut-être plus indiquée) et avons effectué les regroupements en arbres selon la méthode de la distance moyenne.

La classification générale que nous proposons repose sur ces principes. Les configurations finales des arbres semblent satisfaisantes, les différences minimales étant annulées. On prend ainsi en compte les marges d'erreurs.

## CONCLUSION

Les résultats d'une classification lexico-statistique, basée sur le vocabulaire dit "fondamental", donnent une image synthétique, facilement visualisable, de l'état actuel des langues sur le plan lexical. L'hypothèse sous-jacente est que le vocabulaire fondamental est peu susceptible d'emprunt. Entre des langues théoriquement apparentées, les différenciations dans cette part du vocabulaire seraient le résultat de l'évolution naturelle des langues, chaque langue étant supposée se développer indépendamment les unes des autres.

Concernant la démarche purement linguistique, nous pensons qu'il y aurait lieu de compléter cette étude lexicale par des données phonologiques, morphologiques et grammaticales, données (peut-être) encore moins susceptibles d'influences par contact. Des calculs analogues pourraient repenser, par exemple, sur une centaine de questions grammaticales, au sens large.

Notre méthode s'applique parfaitement à des calculs de distances entre différents dialectes d'une même langue. Nous l'avons éprouvée avec les 120 mots de l'ALCAM sur les différents dialectes des langues mafa, giziga, zəlgwa, daba, masa, etc. - ce qui a permis de choisir des dialectes de référence pour la suite des opérations -, mais aussi, avec un corpus beaucoup plus vaste - où seules les différences étaient minutieusement analysées - sur la langue mofu-gudur (cf. D. BARRETEAU 1986).

Il est certain que la méthode est beaucoup plus difficile à appliquer lorsque les langues sont très distantes. Dans ce cas, le recours au principe de transitivité, à des reconstructions ou pseudo-reconstructions est tout à fait indiqué pour éviter des erreurs d'interprétation. Nous conseillerions au premier amateur venu de comparer d'emblée le hausa au yedina ou au kera, l'anglais au russe ! Il va

de soi que la démarche doit être progressive, partir des langues manifestement les plus proches pour aborder, seulement à la fin, les langues les plus éloignées.

Ainsi, la critique souvent formulée à l'encontre de la lexico-statistique, qu'on ne peut établir de ressemblance entre des formes qui n'ont que très peu ou aucun point commun, tombe, si l'on admet que l'on procède par étapes successives, que l'on considère les différentes formes attestées à travers la série complète des langues étudiées et que l'on n'ignore pas les formes déjà reconstruites auxquelles nous pouvons rapprocher les formes observées. Pour les langues tchadiques, nous disposons actuellement de reconstructions pour la majeure partie des 100 termes inclus dans notre corpus : cf. P. NEWMAN / R. MA (1966), P. NEWMAN (1977), H. JUNGRAITHMAYR / K. SHIMIZU (1981). Dans l'établissement de nos jugements, le degré minimum de ressemblance, (1) selon notre méthode, sert précisément à englober des cas d'apparemment supposé sans qu'une ressemblance ne soit manifeste au simple vu de la paire d'items à comparer.

En ce qui concerne la validité d'une classification lexico-statistique, on rappellera le fait que les différents arbres classificatoires obtenus selon nos calculs présentent généralement des configurations analogues. Ceci tendrait à prouver qu'une comparaison "superficielle" ou "profonde" donne des résultats comparables si l'on ne s'en tient pas aux chiffres absolus. Partant, notre hypothèse est qu'une classification génétique - qui reste à élaborer dans le détail - ne serait pas fondamentalement différente de celle que nous proposons.

Nous sommes toutefois bien conscient du fait que seules des reconstructions fondées sur des lois de correspondance peuvent démontrer que des formes relèvent effectivement d'un même étymon, que des langues sont apparentées. La lexico-statistique donne seulement une première estimation sur la ressemblance actuelle des langues entre elles et sur la possibilité de leur apparentement.

Les ambitions de la lexico-statistique ne sont donc pas démesurées - contrairement aux critiques qui lui sont adressées ! Ce n'est qu'une première approche, pratique et rapide, - inutile pour ses détracteurs ! - dans la comparaison des langues en attendant que des recherches plus approfondies viennent confirmer ou infirmer les résultats acquis.

Nous avancerons même que la lexico-statistique a encore de beaux jours devant elle. Dans la perspective d'un programme pluridisciplinaire de recherches comparatives et historiques dans le Bassin du Lac Tchad, il nous semble que la méthode pourrait aussi bien s'adapter à des questions ethnographiques ou géographiques, celles-ci restant à définir. La mise au point de méthodes informatisées ouvre encore de nouvelles perspectives (cf. M. DIEU et al. 1983).

Il y aurait alors grand avantage à comparer les résultats acquis dans les différentes disciplines. En effet, si l'on examine les premiers résultats de recherches sur l'architecture, les instruments aratoires, la pêche, ou certaines institutions comme l'initiation masculine, la chefferie, on se rend compte que nombre de clivages ne correspondent pas à nos regroupements linguistiques, basés, jusqu'à présent, sur le seul vocabulaire fondamental, notre classification n'apportant, semble-t-il, que des informations générales sur les états très anciens des langues, que sur leur évolution "naturelle".

Il semble donc que les linguistes gagneraient à envisager maintenant de nouvelles perspectives de recherche, plus élargies, portant spécialement sur le vocabulaire culturel. L'objectif serait alors de lever le voile sur les échanges et influences historiques dans différents domaines, sur les mouvements migratoires des hommes et des cultures dans cette région, et d'arriver, progressivement, à des hypothèses dépassant les simples données de la tradition orale. Nous prendrions alors en compte ces phénomènes de contact qui ont certainement joué un grand rôle, à toute époque, dans l'évolution des langues et des cultures.



## REFERENCES BIBLIOGRAPHIQUES

- BARRETEAU D. - 1986 - *Description du mofu-gudur (langue de la famille tchadique parlée au Cameroun) : 1. Phonologie, 2. Lexique* - Paris/Yaoundé : ORSTOM/MESRES - 2 vol. à paraître.
- DIEU M. et al. - 1983 - *Atlas Linguistique de l'Afrique Centrale (ALAC) : Structures et méthodes* - Paris/Yaoundé : ACCT/CERDOTOLA.
- DIEU M., P. RENAUD (dir.) - 1983 - *Situation linguistique en Afrique centrale. Inventaire préliminaire : Le Cameroun* - Paris/Yaoundé : ACCT/CERDOTOLA.
- JUNGRAITHMAYR H., K. SHIMIZU - 1981 - *Chadic lexical roots : A first evaluation of the Marburg Chadic Word Catalogue. Vol. II : Tentative reconstruction, grading and distribution* - Berlin : D. Reimer - Marburger Studien zur Afrika- und Asienkunde A.26.
- NEWMAN P. - 1977 - "Chadic classification and reconstructions" - *Afro-asiatic Linguistics* 5 (1).
- NEWMAN P., R. MA - 1966 - "Comparative Chadic : Phonology and lexicon" - *Journal of African Languages* 5 - pp.218-251.

## ANNEXES

En annexes, nous donnerons quelques documents illustrant la méthode développée ici ainsi que quelques résultats :

1. Listes comparatives pour les mots "bouche" et "oeil"
2. Les jugements de ressemblance dans le groupe masa avec quelques exemples illustrant les cinq degrés de ressemblances
3. Les matrices de ressemblance lexicale du groupe wandala avec les positions schématiques des langues wandala, parəkwa et gəvoko
4. Les arbres classificatoires du groupe wandala
5. Les arbres classificatoires simplifiés (taux pondérés ramenés aux dizaines les plus proches, distance moyenne) montrant les relations des langues à l'intérieur de leurs groupes respectifs
6. Classification des langues et dialectes
7. Cartes de localisation des langues

# 1. LISTES COMPARATIVES

## 1.1. bouche

hausa :	baa(kii)	psikye :	me
kera :	(k)u	hya :	fuŋe
zumaya :	vun(aŋu)	bana :	mi
masa :	vun(na)	tsuvan :	ma
musey :	vun(na)	sharwa :	ma
lame :	vun	zizilivøken :	ma(n)
		jimjimøn :	ma(n)
wandala :	we	gufe :	ma
gølvaxdaxa :	ghaya	njanyi :	ma
parøkwa :	vurw(a)	gbwata :	me
gøvoko :	we	daba :	ma
xødi :	wi	besleri :	ma
mabas :	uwi	mbødam :	pakwam / ma
		gavar :	ma
mafa :	ma	buwal :	ma
mefeŋe :	ma	kaða :	ma
cuvok :	ma		
mofu-sud :	mey	munjuk :	(me)me
mofu-nord :	mey		
giziga-sud :	mi	majøra :	me
giziga-nord :	me	jina :	mi
baldamu :	me		
merey :	bazlam	lagwan :	(kal)ge
dugwor :	mey	mser :	ge
zølgwa :	bazlam / aham	afaðe :	ga
gaðuwa :	tepeshe / aham	maslam :	ga
mølokwo :	ma	malgbe :	ga
maða :	(møm)ma	mpaðe :	ga
muyang :	ma	yedina :	gay
wuzlam :	ma		
matal :	gay		
mbuko :	am		
pølasla :	me		

## 1.2. oeil

hausa :	idoo	psikye :	nsə
kera :	dər	hya :	ntse
zumaya :	aray	bana :	mətse
masa :	'ir(a) / ir(a)	tsuvan :	adi(ne)
musey :	id(a)	sharwa :	dye
lame :	'i	zizilivəkən :	ji(n)
wandala :	iyce	jimjimən :	zhi(n)
gəlvaxdaxa :	diya	gude :	gi(n)
parəkwa :	di	njanyi :	dish
gevoko :	odo	gbwata :	di(ce)
xədi :	iri	daba :	həra
mabas :	ili	besleri :	nje
mafa :	day	mbədam :	nje
mefeke :	de	gavar :	nja
cuvok :	ndey	buwal :	nje
mofu-sud :	dey	kada :	ara
mofu-nord :	dey	munjuk :	aray
giziga-sud :	(hi)ri	majəra :	ade
giziga-nord :	re	jina :	iri
baldamu :	aray	lagwan :	sə (sg.) / 'al (pl.)
merey :	dərey	msər :	sə
dugwor :	re	afadə :	tsə
zəlgwa :	ere	maslam :	sə
gađuwa :	ere	malgbe :	sə (sg.) / 'el (pl.)
melokwo :	ele	mpadə :	sə
mada :	iyre	yedina :	yil
muyang :	eri		
wuzlam :	aray		
matal :	yawdi		
mbuko :	ide		
pəlasla :	aray		

## 2. JUGEMENTS DE RESSEMBLANCE DANS LE GROUPE MASA

### 2.1. masa

#### - musey

- 5 x 45 : 1-3-5-6-7-8-13-15-18-20-26-28-29-31-34-35-45-46-47-51-54-55-  
56-58-59-61-64-70-75-76-81-83-84-89-92-97-99-100-101-106-107-  
109-114-118-119
- 4 x 22 : 2-19-30-32-36-40-41-47-49-52-53-57-63-69-71-73-74-98-104-105-  
112
- 3 x 5 : 9-24-50-93-113
- 2 x 9 : 4-10-17-38-82-85-88-102-110
- 1 x 2 : 33-67
- 0 x 17 : 11-12-16-21-25-39-42-43-90-92-95-96-103-111-116-117-120

#### - lame

- 5 x 16 : 1-7-8-31-35-46-51-54-57-64-75-76-99-100-103-119
- 4 x 18 : 5-18-21-24-26-34-41-47-49-50-55-63-69-70-83-84-104-107
- 3 x 14 : 2-11-17-19-20-36-38-45-74-92-101-109-110-112
- 2 x 10 : 6-9-29-40-61-85-88-91-102-106-113
- 1 x 5 : 4-43-56-67-117
- 0 x 36 : 3-10-12-13-15-16-25-28-30-32-33-39-42-47-52-53-58-59-71-73-  
81-82-89-90-92-93-95-96-97-98-105-111-114-116-118-120

### 2.2. musey

#### - lame

- 5 x 16 : 1-7-8-31-35-46-51-54-64-75-76-85-99-100-104-119
- 4 x 19 : 5-18-24-25-26-34-41-47-49-52-55-57-63-70-74-84-102-107-112
- 3 x 15 : 2-16-20-36-42-45-50-69-82-88-91-92-96-101-109
- 2 x 13 : 6-10-17-19-29-32-40-61-67-90-106-110-113
- 1 x 6 : 9-38-47-56-92-117
- 0 x 31 : 3-4-11-12-13-15-21-28-30-32-39-43-53-58-59-71-73-81-83-89-  
93-95-97-98-103-105-111-114-116-118-120

### 2.3. Exemples d'attribution de coefficients de ressemblance entre les langues masa et musey

	masa	/	musey	
Coefficient 5 :	vun		vun	1. bouche
	yam		yam	3. tête
	sii		sii	5. dent
	busuu		busu	18. sang
Coefficient 4 :	ʔir		id	2. oeil
	zumur		sumwu	19. urine
	gor		gwor	30. enfant
	law		lo	32. ciel
Coefficient 3 :	dewel		del	9. cou
	miyok		mek	24. corne
	kew		ngew	50. couteau
Coefficient 2 :	ngus		(tu)mus	4. poil
	míir		mbíi	10. sein
	ruwat		lorot	17. intestins
Coefficient 1 :	henje		jeke	33. nuit
	soʔol		sadii	67. racine
Coefficient 0 :	ɓam		ko	11. main
	terek		yam-cit	(?) 12. ongle
	hif		fuk	16. nombril

### 3. MATRICES DE RESSEMBLANCE LEXICALE : GROUPE WANDALA

<i>Taux minimal</i>	wandala	gɛlvaxdaxa	parəkwa	gəvoko	xɛdi	mabas
wandala	-	35	29	27	24	23
gɛlvaxdaxa	35	-	27	30	21	18
parəkwa	29	27	-	34	19	16
gəvoko	27	30	34	-	34	27
xɛdi	24	21	19	34	-	72
mabas	23	18	16	27	72	-

#### *Taux pondéré*

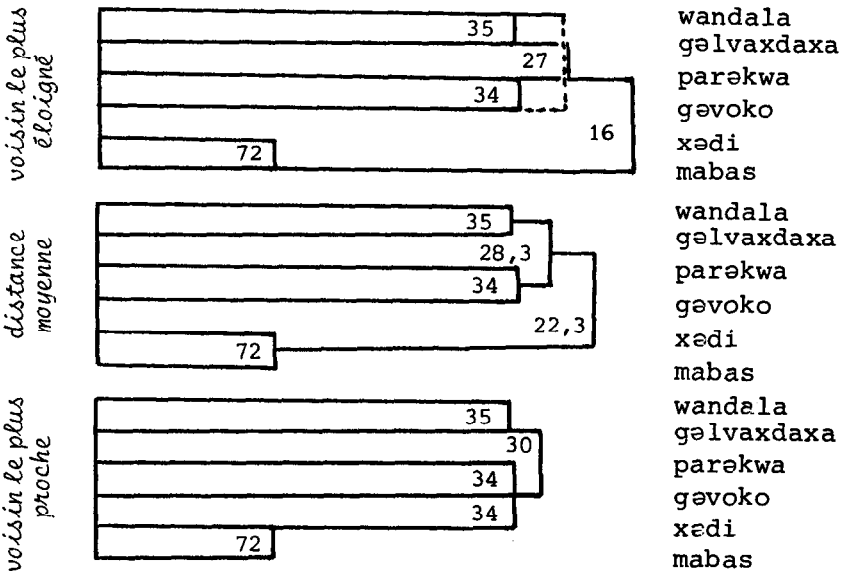
wandala	-	48	41,4	39,8	37,2	36,2
gɛlvaxdaxa	48	-	42,8	42	34	32,8
parəkwa	41,4	42,8	-	44,6	32,6	29,6
gəvoko	39,8	42	44,6	-	49,2	43
xɛdi	37,2	34	32,6	49,2	-	74
mabas	36,2	32,8	29,6	43	74	-

#### *Taux maximal*

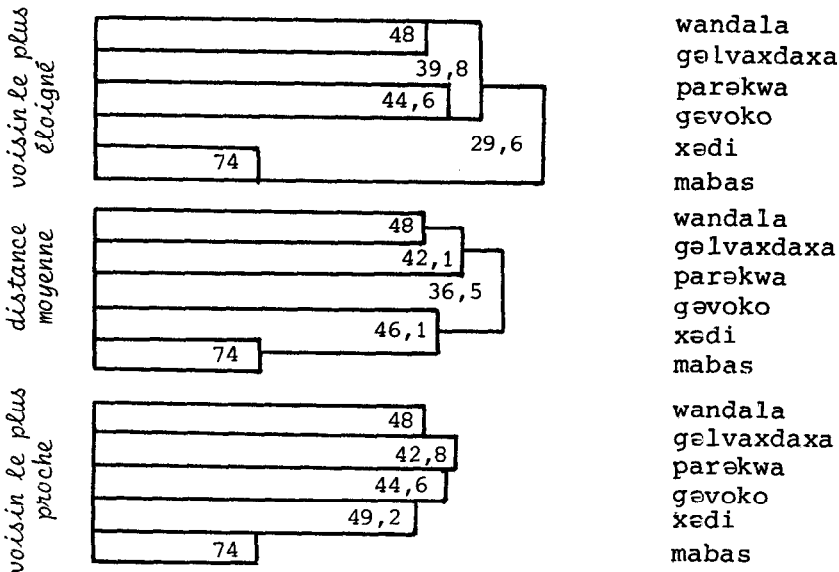
wandala	-	71	68	64	60	58
gɛlvaxdaxa	71	-	66	66	60	55
parəkwa	68	66	-	66	55	52
gəvoko	64	66	66	-	75	68
xɛdi	60	60	55	75	-	87
mabas	58	55	52	68	87	-

#### 4. ARBRES CLASSIFICATOIRES : GROUPE WANDALA

##### 4.1. Taux minimal

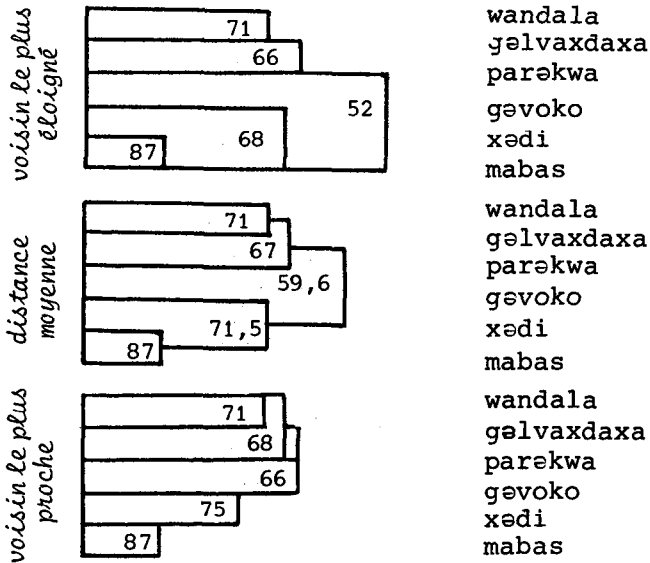


##### 4.2. Taux pondéré

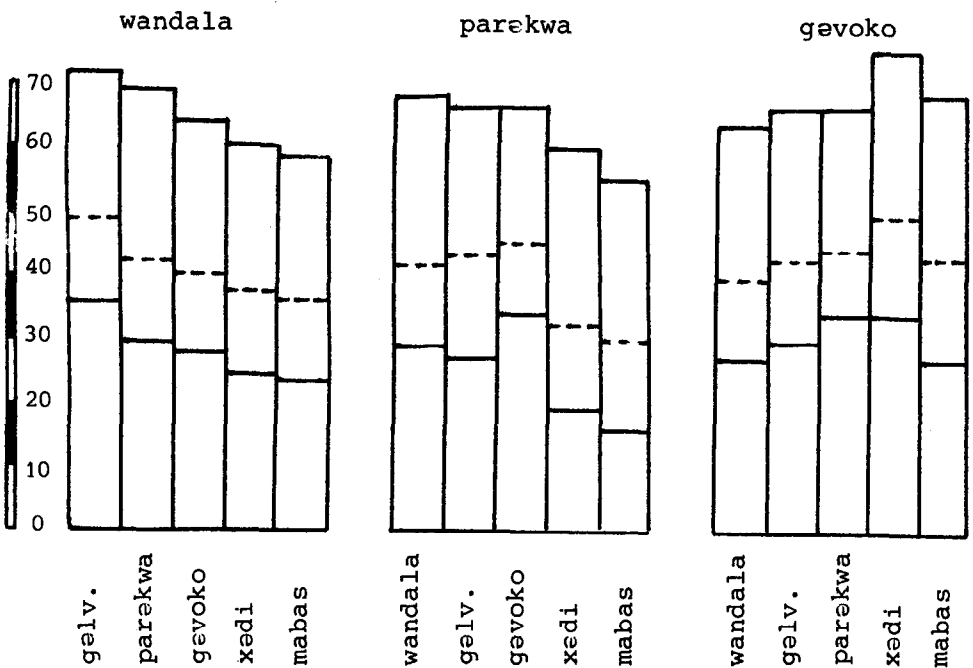




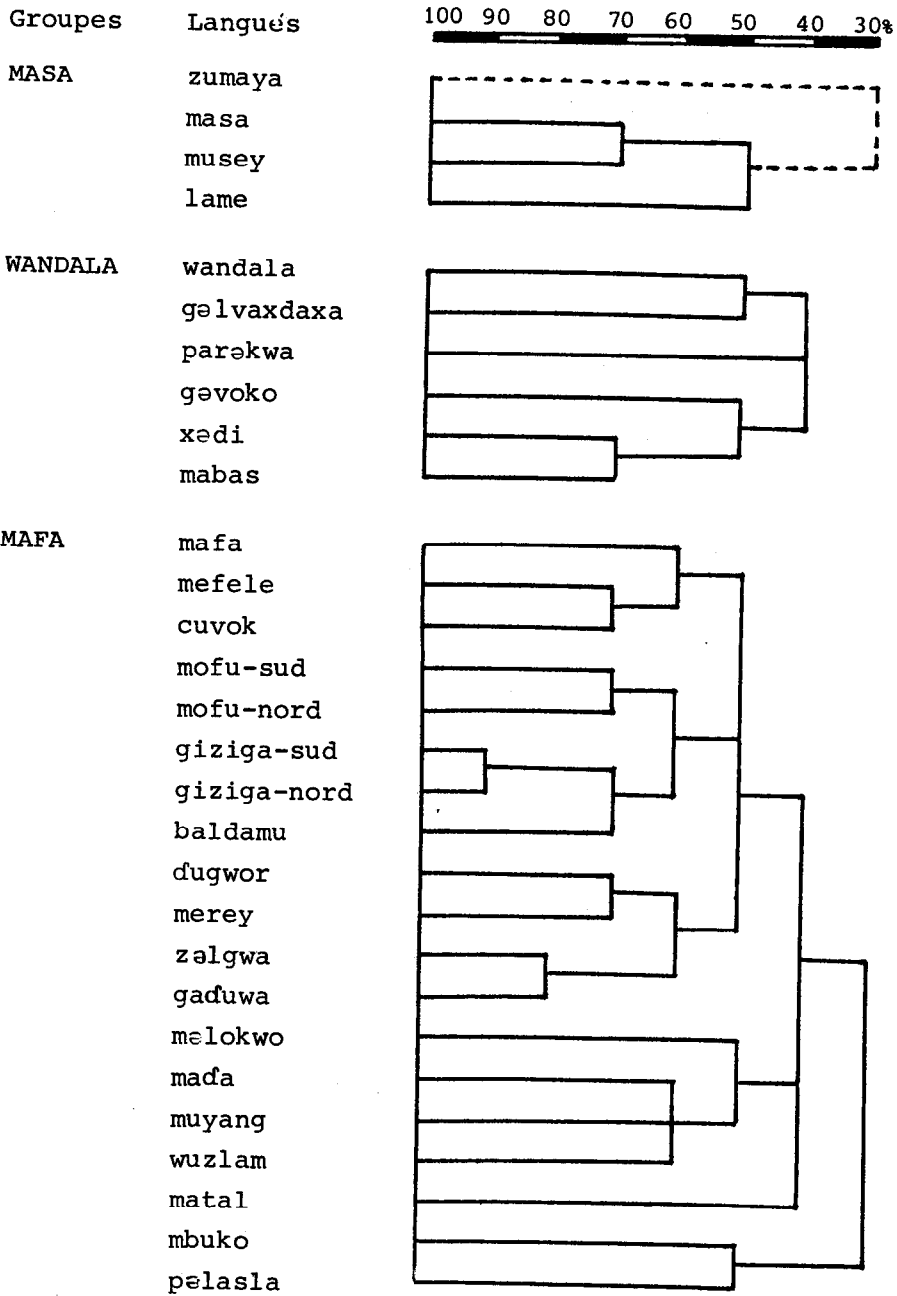
#### 4.3. Taux maximal

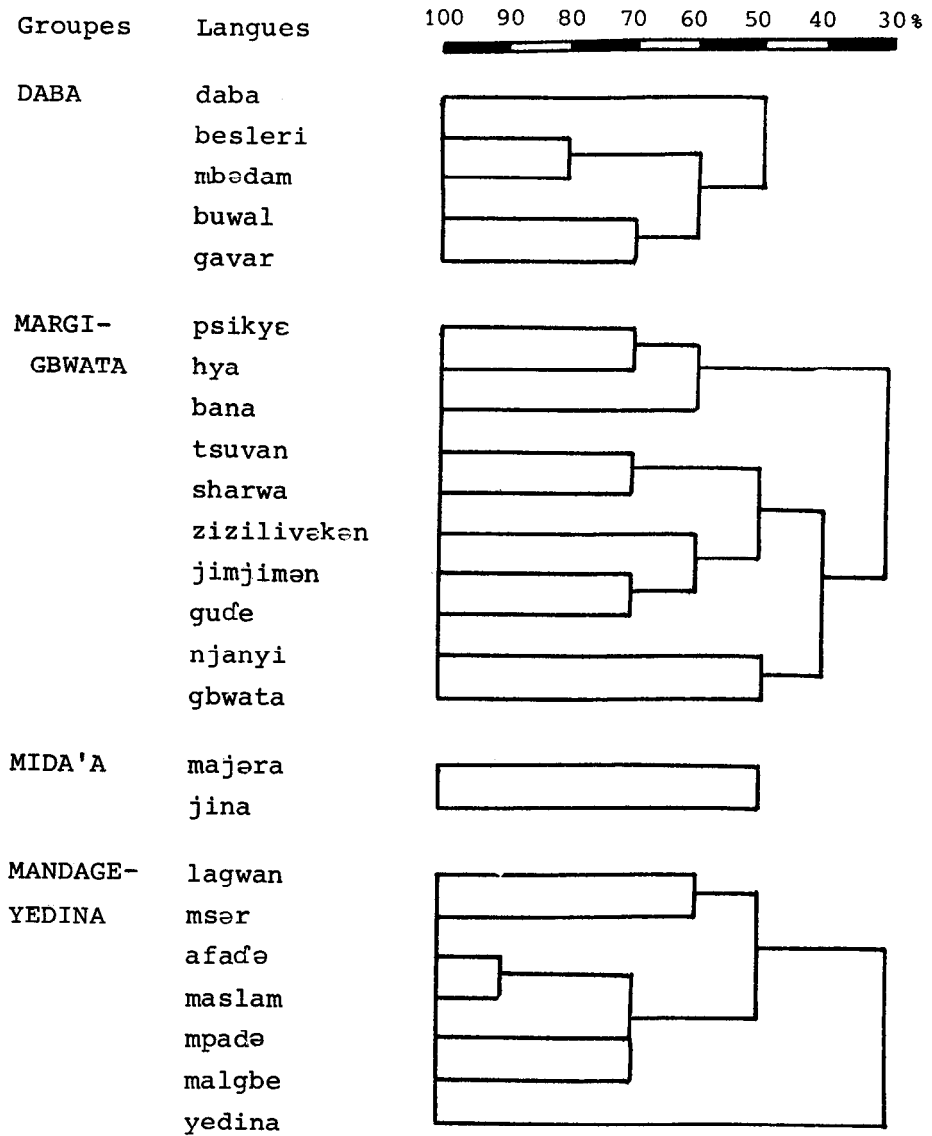


#### 4.4. Positions des langues wandala, parəkwa et gəvoko



# 5. ARBRES CLASSIFICATOIRES SIMPLIFIES





Langues non-classées (langues uniques dans leursgroupes) :

hausa  
 munjuk  
 kada  
 kera

## 6. CLASSIFICATION DES LANGUES ET DIALECTES

Les dialectes d'une même langue seront cités entre parenthèses. Pour informations complémentaires, on se reportera aux cartes ci-joints ainsi qu'aux commentaires sur les noms de langues et de dialectes dans l'Atlas linguistique du Cameroun (DIEU et RENAUD dir. : 1983).

Nous apportons quelques innovations par rapport à l'inventaire et à la classification proposés dans l'Atlas linguistique du Cameroun.

Les rapports entre les groupes n'ayant pas été calculés, la classification restera provisoire dans ce domaine. Par hypothèse, nous distinguons trois sous-branches dans la branche centrale : la sous-branche A reste quasiment inchangée par rapport à la classification de P. NEWMAN (1977). Dans la sous-branche B, nous incluons les groupes kada, masa, munjuk, mida'a, mandage et yedina. C'est une hypothèse qui demanderait à être contrôlée.

### BRANCHE OCCIDENTALE

#### Sous-branche A

##### 1. Groupe hausa

###### 1. hausa

### BRANCHE CENTRALE

#### Sous-branche A

##### 2/3. Groupe combiné wandala/mafa

##### 2. Groupe wandala

###### Sous-groupe nord

###### 2. wandala (wandala, mura, malgwa)

###### 3. gelvaxdaxa (= glayda)

###### Sous-groupe centre

###### 4. parəkwa (= podokwo)

Sous-groupe sud

- (a) 5. gəvoko
- (b) 6. xədi (= hidé)
- 7. mabas

3. Groupe mafa

Sous-groupe sud

- (a) - 8. mafa (mafa-ouest, mafa-centre, mafa-est)
- 9. mefe (mefe, sarak, muhur ; shügule)
- 10. cuvok
- (b) - 11. mofu-sud (Gudal, Mokong, Diméo)
- 12. mofu-nord (Duvangar, Durum-Wazang)
- 13. giziga-sud (Lulu, Muturwa ; Mijivin)
- 14. giziga-nord
- 15. baldamu
- (c) - 16. dugwor (dugwor, mikiri)
- 17. merey (= mofu de Méri)
- 18. zəlgwa (zəlgwa, minew ; gemzek)
- 19. gaduwa
- (d) - 20. məlokwo
- 21. mada
- 22. muyang
- 23. wuzlam (= ouldémé)

Sous-groupe nord-ouest

- 24. matal (= mouktélé)

Sous-groupe nord-est

- 25. mbuko
- 26. pəlasla (pəlasla, ndreme, mbərem, dəmwa, hurza)

4/5. Groupe combiné margi/gbwata

4. Groupe margi

- 27. psikye (psikye, zləŋə, wula) (= kapsiki)
- 28. hya
- 29. bana

5. Groupe gbwata

Sous-groupe nord

- (a) - 30. jimjimən (= jimi)
- 31. gude
- 32. zizilivəkən (= "fali de Jilvu")
- (b) 33. sharwa
- 34. tsuvan

Sous-groupe sud

- 35. njanyi
- 36. gbwata (= bata)

6. Groupe daba

Sous-groupe sud

- 37. daba (daba, mazagway, tpala)

Sous-groupe nord

- (a) 38. besleri (= hina)
- 39. mbədam
- (b) 40. gavar
- 41. buwal (= gadala)

Sous-branche B

7. Groupe kaḏa

- 42. kaḏa (= gidar)

8. Groupe masa

Sous-groupe sud

- (a) 43. masa (masa-ouest, masa-centre, masa-est)
- 44. musey
- (b) 45. lame

Sous-groupe nord

- 46. zumaya

9. Groupe munjuk

- 47. munjuk (muzuk, mpus, beege)
- (cf. vulum, mbara, muskum au Tchad)

10. Groupe mida'a (= kotoko-sud)

- 48. jina (jina, muxule)
- 49. majəra (majəra, kajire-ḏulo, hwaləm)

11/12. Groupe combiné mandage/yedina

11. Groupe mandage (= kotoko-nord)

Sous-groupe sud

50. lagwan (= kotoko de Logone-Birni)

51. msər (= kotoko de Kousséri)

Sous-groupe sud

- 52. afadə

53. maslam (maslam, sahu)

- 54. mpadə (= kotoko de Makari)

- 55. malgbe (= kotoko de Gulfey)

12. Groupe yedina

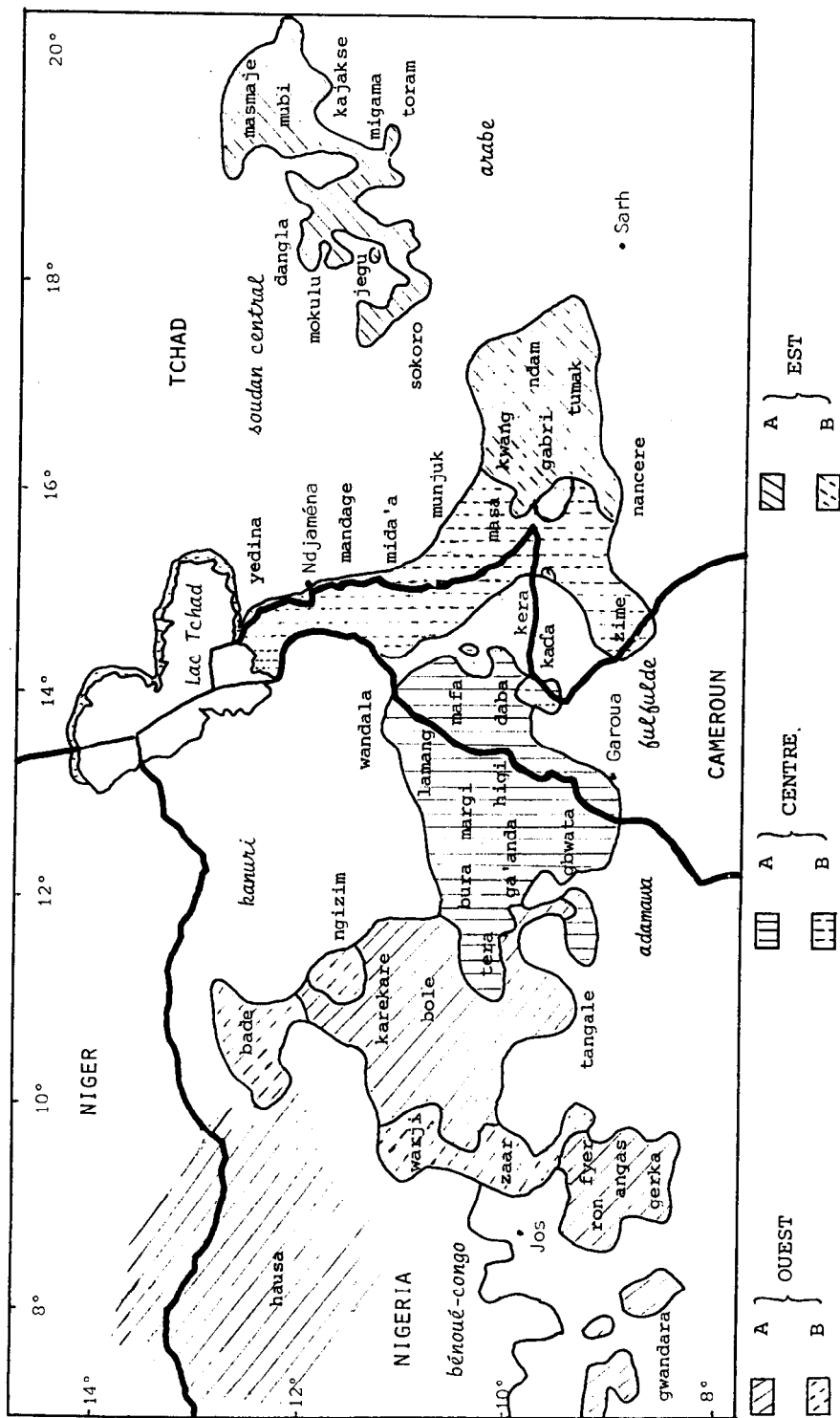
56. yedina (= boudouma)

BRANCHE ORIENTALE

Sous-branche A

Groupe kwang

57. kera



1. SITUATION GENERALE DES LANGUES DE LA FAMILLE TCHADIQUE

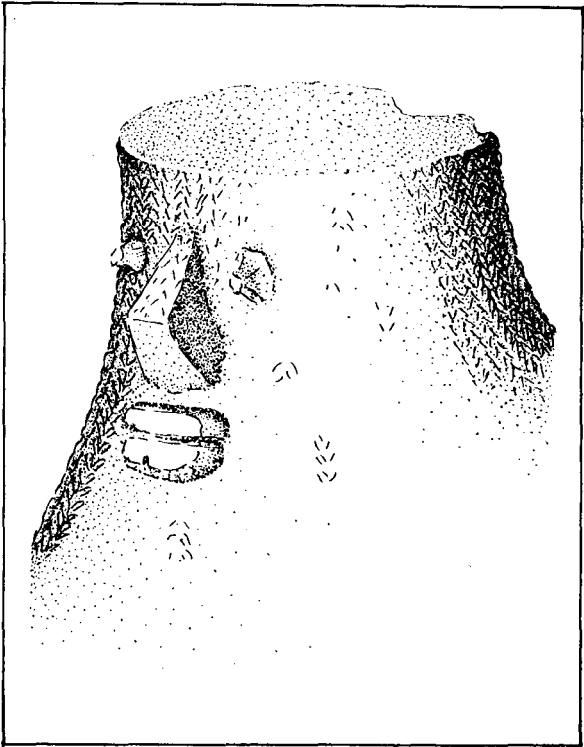




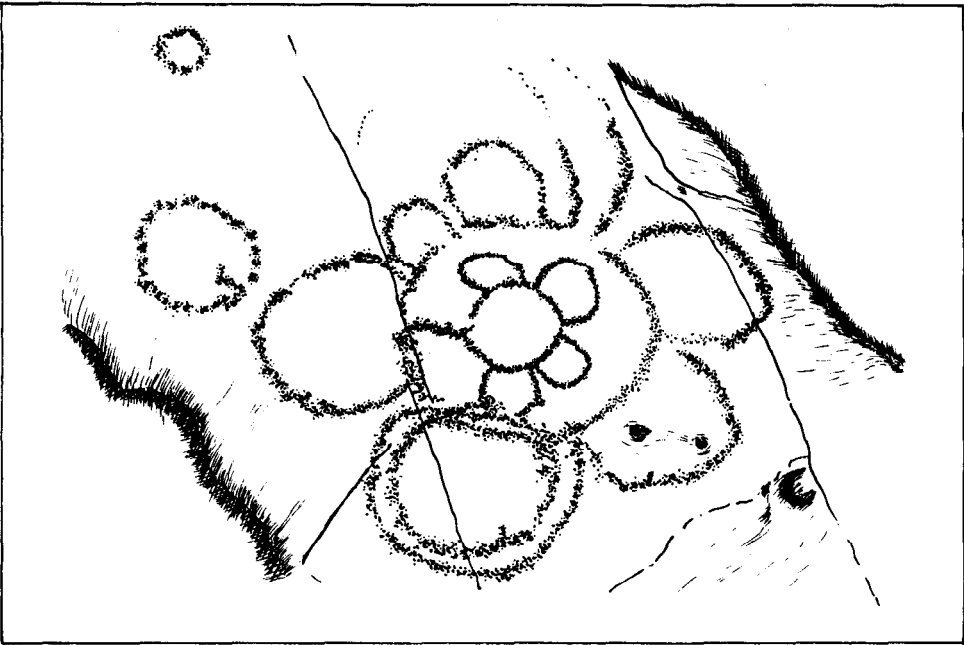
## 2. LES MONTS MANDARA







5 . Poterie sao (Logone-Birni)



6 . Pétroglyphe de Bidzar