

Remarques concernant le classement génétique du tchamba daka (*sámá mūm*)

Raymond Boyd
Laboratoire Ligérien de Linguistique
Université d'Orléans
{version juin 2024}

Le *sámá mūm* 'langue tchamba', que nous appellerons ici simplement le sama, apparaissait sous le nom de « daka » comme unique composant du troisième groupe des langues « Adamaoua » du classement de Greenberg (1966 : 9¹). Actuellement, des deux principales institutions qui s'occupent de la dénomination « standardisée » des langues, seul le Max Planck Institute for Evolutionary Anthropology à travers le site glottolog.org accepte – à ma demande – cette appellation. Le Summer Institute of Linguistics à travers le site ethnologue.com utilise en revanche la désignation « samba daka » pour cette langue. Il s'agit là d'une dénomination de linguiste formulée afin de distinguer cette langue du « samba leko », deux langues dont les locuteurs partagent une identité ethnique (cf. Fardon 1988) que les locuteurs du *sámá mūm* appellent *sámá* et les locuteurs du leko appellent *sámá*². Les « glottocodes » employés par glottolog.org pour abrégier ces deux noms de langues sont : samb1311 (daka) et samb1305 (leko) ; les codes (ou « codets ») ISO³ 639-3 correspondants utilisés par ethnologue.com sont respectivement *ccg* et *ndi*⁴.

Par le passé, le nom « (t)chamba » a souvent été employé pour désigner l'une ou l'autre de ces deux langues ainsi que l'ethnie qui les parle. Cette dénomination se trouve dans la publication contenant les plus anciennes listes de mots que nous possédons pour ces langues (Strümpell 1910) et G-LA l'emploie pour nommer une autre langue du groupe Adamaoua 2 qui ne peut en principe pas être différente du « lekon ». Ceci reflète l'usage local : « tchamba » est en effet l'hétéronyme généralement employé pour désigner l'ethnie qui nous concerne et « tchamba daka » et « tchamba leko » sont des désignations possibles pour les deux langues qu'elle parle. En tout cas, ce sont ces termes que les Sama utilisent en parlant une autre langue que la leur pour s'autodésigner.

L'origine de la forme « tchamba » demeure inconnue mais on pourrait raisonnablement envisager l'hypothèse que les Peuls, dominants dans la région, avaient formé un pluriel régulier de

¹ Cet ouvrage sera référencé désormais par le sigle G-LA.

² On a ainsi *sám(bá) wāŋ(á)* 'langue tchamba', vd. Fabre 2003. Il s'agit du « lekon » de G-LA qui se trouve dans le deuxième groupe des langues Adamaoua. Nous confirmons d'ailleurs ce que le lecteur a bien pu deviner : le sama n'a pas de consonnes minasales dans son inventaire phonologique natif et le leko n'en emploie pas en position finale.

³ ISO est l'acronyme de la International Organization for Standardization. De façon aberrante, l'accès aux normes publiées par cet organisme est payant et par conséquent elles ne sont pas employées en dehors de milieux techniques fermés. Le site glottolog.org les rappelle pour mémoire.

⁴ Dans la suite de cet exposé, nous rapporterons lors de leur première citation les codes attribués aux langues par les deux sites et aux familles de langues par Glottolog.

samba et que c'est cette forme qui s'est généralisée. Cette association du nom à minasale intervocalique à la langue dominante renforce une aura pré-existante : à l'origine un nom de patriclan chez les locuteurs du leko, *samba* avait acquis une connotation de prestige en raison d'une pré-éminence accordée aux Samba dans le domaine rituel. De plus, sur la base de la consonne initiale, les Sama de confession musulmane voient un rapport étymologique entre l'ethnonyme tchamba et le Djebel Chams⁵, plus haut pic du Sultanat d'Oman, d'où ils auraient émigré⁶.

Quant au nom daka (*dākā*), il n'est employé seul (en dehors d'une littérature académique) que pour désigner une ethnie établie dans le sud-ouest du grand ensemble tchamba. Comme dans le cas du terme *samba*, il pourrait s'agir du nom d'un patriclan qui se transforme en ethnonyme lorsque d'autres groupes lui accordent une pré-éminence et en adoptent l'identité, indépendamment du lignage⁷.

Les deux sites, Glottolog et Ethnologue, rejettent le classement de G-LA et rattachent le sama génétiquement à l'ensemble « Bénoué-Congo⁸ » (BC). L'objectif de l'exposé qui suit est de soutenir que ce changement d'affiliation est injustifié. Pour cela, je résume d'abord (§1 à 5) trois méthodologies auxquelles on peut avoir recours lorsqu'il s'agit d'établir un classement « génétique » des langues, c'est-à-dire, un classement qui regroupe les langues selon le fait d'avoir ou non un ancêtre commun présumé. Je m'exprimerai dans ces paragraphes comme si je m'adressais à des débutants, non pas parce que je pense m'adresser à des profanes mais parce que je pense que, si l'on accepte mes principes de base, on ne pourra pas conclure à ce reclassement et il faut par conséquent les expliciter. Je tirerai ensuite les conclusions qui s'imposent (§6-7), en particulier par rapport à l'étude qui a annoncé le reclassement que je conteste (Bennett 1983), et je considérerai les méandres que cette proposition a connus par la suite (§8). Je proposerai alors qu'on reprenne le classement des langues africaines dans une optique différente (§9) dont le point de départ sera toujours G-LA. Pour cela, il me faudra relier ce point aux développements postérieurs (§10-12) afin d'esquisser une voie parallèle à celle qui a été définie en §7-8.

⁵ Il s'agit en fait d'un ش initial, phonétiquement [ʃ].

⁶ Remarquons toutefois que G-LA cite également une langue « Cham » dans le groupe Adamaoua 1. D'après Kleinwillinghöfer (2014), il s'agit d'un hétéronyme qui désigne les locuteurs de deux dialectes parlés par deux populations qui manifestent une différence ethnique non précisée et qui n'emploient aucun autonyme commun. Ce nom désigne aujourd'hui les villages habités par ces populations, Cham-Mwona et Cham-Kindiyo.

⁷ Le sama connaît un certain degré de variation dialectale non encore décrit avec précision mais suffisant pour avoir généré une *lingua franca*, appelée *n̄-nàk-é-nyāā-rè* (2s-faire-aspect-comment-interrogatif 'comment va-tu ?'), utilisée lorsque différents groupes se rencontrent. L'onomastique est compliquée dans cette région et parmi les désignations recueillies des divers dialectes, on ne peut savoir s'il s'agit de noms de langues, d'ethnies, de clans ou de lignages.

⁸ Glottolog réaménage le classement de Greenberg en prévoyant une famille Atlantique-Congo (glottocode atla1278), plus réduite que l'ancien Niger-Congo, avec une branche Volta-Congo (volt1241). Sur celle-ci se trouvent le Bénoué-Congo (benu1247) et le Volta-Congo nord (nort3149). Ce dernier regroupe sur des sous-branches différentes les langues gour (on écrira « gur » ici) et Adamaoua qui étaient chez G-LA des groupes coordonnés du Niger-Congo.

Cette longue discussion nous montrera que l'existence d'une dualité d'avis sur ce sujet est basée sur une opposition de points de vue concernant la comparaison multilatérale. Dans les deux dernières sections, j'examinerai le contexte où ce débat a lieu et je m'y situerai (§12). Ensuite (§13) je m'interrogerai sur les raisons pour lesquelles les deux positions semblent irréconciliables.

1. Concernant les façons d'aborder la diachronie des langues

Afin de connaître l'évolution historique des langues, on aura recours, d'un point de vue théorique, à l'une ou à l'autre de trois méthodologies :

- celle de la comparaison des langues en grand nombre (*mass comparison*), appelée aussi *comparaison multilatérale*, mise en œuvre pour les langues africaines par G-LA et servant à établir un classement appelé génétique⁹ ;
- celle de la linguistique historique ou comparative, pratiquée selon diverses approches faisant partie d'une tradition qui a son origine chez les Néogrammariens du 19^{ème} siècle et servant à préciser les traits d'une supposée protolanguage, ancêtre unique des langues observées ;
- et celle de la lexicostatistique, utilisée par Swadesh dans le cadre de la glottochronologie (voir sa bibliographie : https://es.wikipedia.org/wiki/Morris_Swadesh#Bibliografía), mais qui sert également à déterminer les contours de communautés linguistiques contemporaines.

a) La *comparaison multilatérale* repose sur le principe que, plus le nombre de langues prises en compte est grand, plus leur classement génétique dans de grands ensembles (ayant une origine historique commune) devient clair. Ainsi dans la pratique, la simple inspection du vocabulaire de base des langues parlées sur le plateau de l'Adamaoua et dans ses alentours nous permet d'identifier d'emblée l'existence de groupes de langues manifestement apparentées - par exemple, le groupe connu maintenant sous l'appellation « samba-dourou¹⁰ » composé du dii¹¹, du tchamba leko et de plusieurs autres langues avoisinantes (en gros, les groupes Adamaoua 2 et 4 de G-LA). Lorsqu'on met en vis-à-vis de plus en plus de groupes de ce type, on voit apparaître des agglomérations auxquelles on arrive à rattacher d'autres langues plus isolées linguistiquement, en particulier lorsqu'on porte un regard sur des traits qui dépassent le simple niveau lexical. Ainsi on verra d'abord que le groupe samba-dourou a, malgré des différences importantes, plus de traits en commun avec le « mboumique¹² », l'ancien groupe Adamaoua 6 formé par le mboum¹³ et ses parents à lui, qu'avec d'autres groupes pourtant géographiquement proches comme ceux des langues classées « jukunoïdes¹⁴ » ou des langues « (BC) Plateau¹⁵ » qui seront apparentés d'abord entre eux. Ensuite, nous pouvons constater que des isolats - langues ou petits groupes peu différenciés comme le « bena-

⁹ On verra toutefois dans la suite de cet exposé l'importance qu'il faut accorder à l'opposition « généalogique / génétique » introduite par Manessy (1992).

¹⁰ Glottocode samb1323.

¹¹ Glottocode diii1241, ISO 639-3 dur.

¹² Glottocode mbum1257.

¹³ Glottocode mbum1254, ISO 639-3 mdd.

¹⁴ Glottocode juku1257.

¹⁵ Glottocode benu1248.

mboi¹⁶ » (bɛna-mboi) - l'ancien yungur qui constituait le groupe Adamaoua 7 de G-LA - se rapprochent davantage des groupes samba dourou et mboumique que de tout autre grand ensemble. Ce processus d'agglomération passe par la constitution d'ensembles de plus en plus grands et atteint sa fin lorsqu'on s'aperçoit que les différences entre ensembles deviennent bien plus importantes que les ressemblances et qu'une faille se dessine. C'est là où on trace les frontières entre familles en fonction du postulat d'origines différentes. C'est ainsi qu'on place les langues « Adamaoua » en compagnie des autres branches de la famille que G-LA a dénommée « Niger-Congo » alors que les langues tchadiques, quoique souvent en contact avec les langues Adamaoua, ont très peu de traits en commun avec elles mais présentent en revanche des ressemblances importantes avec d'autres langues et groupes de langues qui vont constituer la famille « afroasiatique ».

b) Moyennant la méthode *comparative* ou *historique*, on pourra tenter de remonter aux protolangues de groupes compacts (comme par exemple le samba-dourou ou le mboumique que nous venons de citer) dont on pourra repartir, dans le cas idéal, en les comparant entre elles afin de reconstruire un ancêtre encore plus lointain dans le temps. Toutefois, à l'intérieur des grands ensembles reconnus par la comparaison multilatérale, certaines langues (celles que je qualifie d'isolats relatifs) vont rester à l'écart de cette entreprise de reconstruction faute des maillons historiques permettant de suivre les pas franchis pendant leur dispersion et leur diversification dans le temps.

2. Excursus 1 : la comparaison multilatérale et la linguistique historique contrastées

La tâche de la linguistique historique est de formuler, à la vue des observables, des hypothèses descriptives et explicatives des changements linguistiques qui auraient mené d'un hypothétique ancêtre unique à la multiplicité des parlars actuels. Cette méthode est aux prises avec le déroulement de processus diachroniques concrets et ses résultats évoluent avec l'amélioration progressive des connaissances des langues comparées. Plus on connaît l'architecture et le fonctionnement de ces langues, plus on sera capable de déterminer la nature du système plus ancien postulé et des changements qui l'ont affecté.

La méthode de comparaison multilatérale opère en revanche au niveau de nos connaissances globales des procédés de changement dans les langues naturelles, acquises à travers la pratique de la méthode historique ou, au minimum, de la description de langues apparentées. Elle porte un jugement sur la probabilité que telle langue ait pu provenir d'une langue plus ancienne au même titre que telles autres, sans prétendre accéder à cette protolangue, encore moins aux processus de diversification qui l'ont affectée. Par conséquent, tant que ces connaissances globales n'auront pas considérablement évolué, une fois qu'on aura effectué un classement par comparaison multilatérale, celui-ci est définitif. En effet, dans la mesure où les décisions de classement reposent sur une simple inspection par un observateur averti (qui devient d'ailleurs plus averti au fur et à mesure qu'il agrandit son ensemble de référence), il n'est pas envisageable que des observateurs divergent dans leurs avis, ou si tel était le cas, qu'ils ne puissent pas trancher entre les différentes propositions à la vue d'argumentations pertinentes. On s'en référera comme exemple aux raisonnements de Greenberg

¹⁶ Glottocode bena1258.

(1974) afin de justifier l'inclusion du groupe gbaya dans l'ensemble oubanguien plutôt que dans son voisin Adamaoua. Il faut toutefois admettre en théorie qu'on puisse être incapable de relier une langue déterminée à un grand ensemble quelconque ou que cette liaison apparaisse comme ambiguë, et donc que cette langue soit complètement isolée.

Comme Greenberg l'a fait remarquer à de multiples reprises (notamment dans Greenberg 1957 : 35-45), l'inspection, associée ici à la comparaison multilatérale, précède de fait l'application de la méthode comparative. En effet, pour donner naissance à celle-ci, les Néogrammairiens n'avaient pas fait autre chose qu'observer dans plusieurs langues l'existence de termes étonnamment ressemblants et, encore plus important, le fait que les correspondances entre les composants phonétiques de ces termes semblaient régulières, autrement dit, qu'elles se reproduisaient dans une pluralité de termes comparés. Ils ont ensuite compris que l'explication de ces ressemblances résidait dans des processus de changement phonétique à partir de formes plus anciennes. C'est grâce à près de deux siècles de progrès dans la discipline issue de ces découvertes et au raffinement conséquent de notre perception des liens historiques possibles entre langues que nous nous trouvons aujourd'hui en mesure d'effectuer des classements par comparaison multilatérale.

3. Reprise : la lexicostatistique et ses limites

c) La lexicostatistique opère, quant à elle, strictement dans le cadre de langues dont la parenté est déjà reconnue. Son utilité principale réside dans le traitement de variétés de langues à un niveau de diversification relativement faible. En effet, on peut se trouver face à un ensemble de parlars dont l'arborescence génétique est floue car elle diffère selon le phénomène de changement pris en considération. Si on se focalise alors sur le lexique, une étude statistique peut révéler l'existence de groupes définis selon leur homogénéité à ce niveau, indicative tant de contact prolongé et d'échanges que d'héritage, et parfois annonciatrice de clivages qui vont s'accroître à l'avenir.

Dans le cas d'une diversification plus importante de langues présentant une arborescence claire, la lexicostatistique peut être mise au service de la glottochronologie dans le but de situer les différentes ramifications, soit dans un temps absolu, si le taux d'usure du vocabulaire de base (ou de ses composants) est jugé constant, soit dans un temps relatif dans le cas contraire.

Or, si la comparaison multilatérale peut servir de prélude aux études de linguistique historique en définissant le domaine d'application de celles-ci, la lexicostatistique n'a aucun rôle à jouer au niveau du classement génétique à niveau temporel profond. En particulier, elle ne peut ni se substituer à la comparaison multilatérale pour la définition d'ensembles de langues apparentées ni confirmer ou infirmer ses résultats.

On peut imaginer que la statistique lexicale puisse fournir une sorte d'index de cohérence des ensembles en précisant le degré de divergence entre les langues concernées et cela peut-être au-delà du vocabulaire de base en s'appliquant à différents domaines sémantiques ; mais elle n'a aucune fonction dans la définition de ces ensembles.

En effet, la comparaison multilatérale prend en considération non seulement des ressemblances lexicales, susceptibles d'être comptées, mais aussi des traits qualitatifs, définissables mais non dénombrables, qui caractérisent les langues apparentées. En plus, la lexicostatistique n'a aucune manière de distinguer les ressemblances lexicales dues à l'héritage de celles qui sont une

conséquence des multiples phénomènes de contact qui se manifestent dans toutes les familles linguistiques présentes en Afrique.

Ce dernier écueil est reconnu par Bennett et Sterk (1977, désormais B&S) qui essaient d'effectuer une réorganisation de la famille Niger-Congo à l'aide de la lexicostatistique : « The nature of lexicostatistics is such that geographic and social proximity tends to increase cognacy scores significantly ». Ils prétendent contrer la déviation statistique provoquée par les faits de contact en introduisant une dose d' « innovationnisme » : ils basent leur étude « not only on the statistical studies mentioned, but also on investigations aimed at establishing the distribution of shared lexical and phonologic[al] innovations. Where the two types of study disagreed, the innovation-based evidence was usually given priority ». Or, au niveau théorique, il est difficile d'admettre la validité de l'identification d'innovations en absence de toute reconstruction de protolangue. Au niveau pratique, la recherche d'innovations, en absence de rupture totale de contact entre parlers, s'avère généralement vaine : à chaque fois on relève soit une langue du soi-disant groupe innovateur qui n'a pas fait l'innovation, soit une langue du soi-disant groupe conservateur qui l'a faite, soit les deux¹⁷, car l'osmose entre variantes demeure, sans parler des cas occasionnels d'innovations parallèles indépendantes.

Le dernier facteur qui motive le rejet de l'emploi de la lexicostatistique à des fins de classement génétique à grande échelle est sa subordination absolue à des représentations sous forme d'arborescences qui sont intrinsèquement en contradiction avec les phénomènes omniprésents de contact. Il serait envisageable de définir un ensemble de traits, dont lexicaux, de donner un poids numérique à chacun de ces traits et de projeter les chiffres obtenus sur un espace de façon à représenter la proximité de chaque langue à chacune des autres. Cette représentation pourrait être aménagée en fonction de la structuration de l'ensemble de traits pris en compte et éviterait la difficulté des arborescences à rendre visible les rapports de proximité multiples comme dans le cas des langues toula-wadja appartenant à l'ensemble Adamaoua de la famille Niger-Congo. Le classement de ces langues par G-LA reflète leur situation géographique mais Kleinewillinghöfer (1996) pointent les importantes ressemblances morphologiques avec les langues gur. L'imposition d'une arborescence oblige de les rattacher à l'un ou à l'autre de ces deux ensembles, Adamaoua ou gur ; le choix d'une projection spatiale permettrait de visualiser la réalité des rapprochements multiples.

4. Excursus 2 : les emplois du terme « innovation »

La notion d'innovation devient imprécise lorsqu'elle est employée non seulement dans un cadre lexicostatistique comme celui développé par B&S 1977 mais aussi dans toutes les études qui prétendent aborder les traits d'un supposé proto-Niger-Congo. En linguistique historique, on parle d'innovation lorsque une forme de la protolangue subit une altération à un moment donné. Si cette altération n'est pas générale, des dialectes sont créés et la forme originale toujours présente dans un parler particulier est une rétention¹⁸.

¹⁷ Vd. le cas du tiv évoqué ci-dessous, note 115.

¹⁸ Avec le passage du temps, des changements différents peuvent intervenir dans chaque filiation dialectale de telle façon que la forme originelle n'est conservée nulle part.

Prenons un exemple d'une innovation phonologique hypothétique dans le mboumique, un groupe Adamaoua compact, concernant la racine 'tête' (cf. B&S 1977 : 249, 270) : nous trouvons *tíl* en karang¹⁹, *tól* en moundang²⁰, *tí* en toupouri²¹. Le sens de l'innovation semble clair : la forme d'origine est celle conservée en karang et dans la presque totalité des langues mboumiques occidentales. Cette conclusion est justifiée d'abord par la fréquence du phénomène de l'avancement d'une voyelle postérieure fermée après une consonne alvéolaire dans les langues de la région ainsi que par la présence de la racine originale ailleurs en Adamaoua, notamment dans le groupe boua où elle apparaît avec consonne initiale *s-* en niellim²² et *h-* (< *s-*) en goula Iro²³. Nous pourrions envisager ce changement comme une suite de deux étapes : dans la première, il y aurait une centralisation de la voyelle dans une proto-moundang/toupouri et dans la seconde un passage à une articulation totalement antérieure caractérisant le seul moundang lors d'une nouvelle dialectalisation.

Considérons maintenant les reconstructions proposées pour 'tête' en bantou commun, rassemblées sur https://www.africamuseum.be/fr/research/discover/human_sciences/culture_society/blr²⁴ :

-*tú*, 3/4 (Guthrie 1800)

-*túì*, 3/4 (Guthrie 1811), -*túè* (Guthrie 1808)

cf. -*jòtú*, 3/4 (Guthrie 2177)

On ne peut qu'être intrigué par l'existence d'un autre ensemble de reconstructions ayant le sens 'front, (haut du) visage' :

-*túò*, (?) (Bourquin, Meeusen)

-*ciú*, 14/6 (Guthrie 347), -*ció* (Guthrie 346)

-*cú*, 14/6 (Guthrie 391)

cf. -*jòcú*, 14 (Guthrie 2148)

On signale également l'existence de rapprochements avec -*jícò*, 5/6 (Guthrie 2030), -*jíncò* (Guthrie 2077) 'oeil'.

¹⁹ Glottocode kara1478, ISO 639-3 kzs. C'est la racine choisie par G-LA (11) comme exemple d'une correspondance de classes nominales entre branches du Niger-Congo : « thus we have...the *-le* class in Laka *tu-l* 'head' (cf. Bantu *le-to* 'head', with the same element as a prefix) ». Bien entendu *le-to* serait préférentiellement 'oreille' en bantou commun (les tons de 'tête (3/4)' et d' 'oreille (5/6)' étant par ailleurs différents). C'est ce genre d'erreur qui donne prise aux adversaires de Greenberg alors que le fait intéressant n'est pas la confusion du chercheur mais la réalité de l'instabilité de la classification entre ensembles qui permet cette confusion : à l'instar de l'oti-volta (Manessy 1975 : 122) et contrairement au bantou, le marqueur *-m*, associé aussi au genre 3/4 en bantou, est réservé presque universellement aux non-comptables en Adamaoua.

²⁰ Glottocode mund1325, ISO 639-3 mua.

²¹ Glottocode tupu1244, ISO 639-6 tui.

²² Glottocode niell1243, ISO639-3 nie.

²³ Glottocode gula1265, ISO639-3 glj.

²⁴ Le lecteur pourra consulter ce site pour les références bibliographiques pertinentes.

Tournons-nous maintenant vers l'ensemble gur et en particulier vers les langues Oti-Volta²⁵ (Manessy 1975). Pour 'tête' Manessy reconstruit **z,u* pour une série constituée des formes *zu*, *zu(-g)* et *yu(-l/r)* ou *yi(-l/r)*²⁶ au singulier présentant parfois un schème tonal BH. Le classificateur nominal correspondant serait alors soit **k-*, soit **d-* (= **ɗ-*). En ce qui concerne 'front', Manessy reconstruit à partir d'une série partielle²⁷ **ti/un* avec une consonne radicale finale qu'il appelle « élargissement » en s'interrogeant sur une éventuelle fonction morphologique originelle ; son classificateur au singulier est encore **d-* (= **ɗ-*). Certaines langues préfixent la racine 'tête', *yu/i*, ce qui nous amène à nous interroger au sujet d'un éventuel rapprochement avec les **ji/u-* observés dans les dissyllabes du bantou commun. Manessy rapporte également une racine **ni/un* 'visage'²⁸ qui est presque identique à **ni/un* 'oeil'.

Regardons enfin la situation en Adamaoua, d'abord pour 'tête' : nous trouvons une forme *yu/i(l)*, générale en Adamaoua central²⁹ à l'exception du samba-dourou nord³⁰ où l'on a *zu(l)*³¹, qui nous rappelle les racines présentes en gur³². En revanche, le tula-waja³³, le longuda³⁴ et le bena-mboi³⁵ ont *dur/l* (> *nur* en mboi) qui pourrait être une évolution phonétique à relier aux très rares langues oti-volta qui présentent une affriquée *j-* initiale. Le terme 'front' n'est malheureusement pas disponible dans la plupart des langues Adamaoua mais dans l'Adamaoua central où des lexiques plus importants ont été collectés, on retrouve largement la racine *ti(n)* qui correspond bien à la racine gur **ti/un*. Encore plus remarquable est la présence en samba-leko d'une racine *gbàʔ* 'visage', *gbàsèèn* 'front'³⁶. Il s'agit là d'une forme rapprochable de l'un des concurrents de la racine gur (vd. note 27) et soulève la possibilité que l'autre concurrent *di:(r)* puisse correspondre à Adamaoua *dur/l* 'tête' déjà mentionné. Si nous nous tournons enfin vers la racine 'oeil', nous voyons apparaître de manière quasi générale un correspondant de la racine gur avec la même alternance *i/u*, même dans les langues ayant les lexiques les plus divergents de l'ensemble Adamaoua comme le burak³⁷, le jen³⁸ et le kam³⁹. Or, dans les langues boua nous rencontrons exceptionnellement une racine qui pourrait être un

²⁵ Glottocode otiv1239, toutefois les langues référencées sur glottolog.org ne sont pas identiques à celles qui apparaissent dans les séries de Manessy (1975).

²⁶ Racine également présente en toussian (glottocode tusi1238).

²⁷ Cette racine est concurrencée par *?gbe* et par *?di:(r)*.

²⁸ Cf. sama *nīn* 'front'.

²⁹ Glottocode samb1322. Par ailleurs, sa présence en day (glottocode dayy1236, ISO639-3 dai) est probable.

³⁰ Glottocode nort3259.

³¹ Par ailleurs, en mumuye-yandang (Glottocode mumu1249), en dehors du « mumuyic » (Glottocode mumu1250), *yu* > *u*.

³² La forme *zu* est également présente en gbaya (Glottocode gbay1279).

³³ Glottocode tula1250.

³⁴ Glottocode long1389, ISO639-3 lnu.

³⁵ Glottocode bena1258.

³⁶ Composé possible avec la racine *sen* qui apparaît exceptionnellement en dong (vd. §12.3) avec le sens 'oeil' ayant des correspondances éventuelles en nilo-saharien.

³⁷ Glottocode bura1271, ISO639-3 bys.

³⁸ Glottocode jenn1241.

³⁹ Glottocode kamm1249, ISO639-3 kdx, réputé proche du gur comme le tula-waja et le longuda.

correspondant de *yi(I)* ‘tête’ relevé dans une partie du samba-dourou sud, notamment en « sambaique »⁴⁰.

Notre discussion est partie d’un changement phonétique que nous pouvons expliquer en termes d’un processus connu. Dès que nous avons regardé ailleurs, nous avons rencontré des anomalies comme la présence d’une fricative initiale *z* dans le groupe samba-dourou nord qui ne serait explicable qu’en supposant une inversion du processus $z > y$ en gur. Dans ce cas, l’interprétation est déjà obscurcie par l’absence de la reconstruction d’un système phonologique qui l’aurait encadrée. Elle est rendue encore plus compliquée par la possibilité de changements sémantiques qui nous conduisent à préciser un ensemble de notions qui, par divers phénomènes relevant de la métonymie, peuvent se substituer les unes aux autres. Ceci étant, lorsque nous trouvons en sama la forme *tīī* ‘tête’ ou, encore plus loin, *tī* ‘sur’ en zande⁴¹ oubanguien, que devons-nous en penser ? Elles n’ont certainement aucun rapport avec le changement en mboumique mais pourrait-il s’agir de passages d’un **tu* ancien à *tī* dont le cas mboumique serait un calque ? Ou ne serait-il pas plus simple, dans le cas du sama, de penser au remplacement de la racine d’origine (apparentée peut-être au **zu* présent en gur⁴²) par la racine ‘front’, représentée tant en bantou qu’en gur, moyennant une altération sémantique⁴³ ? En dehors d’un cadre de reconstruction comparant plusieurs langues apparentées de près, la notion d’innovation manque de transparence et perd toute utilité pour le classement génétique.

A l’inverse des innovations phonétiques, les innovations sémantiques ont un intérêt particulier pour la linguistique historique dans la mesure où elles permettent d’élargir le nombre de racines comparables. La recherche de telles innovations est pourtant peu développée, freinée sans doute par l’obligation de restreindre fortement la variation de sens permise lorsqu’on ne travaille que sur des lexiques de base, tant en comparaison multilatérale qu’en lexicostatistique. Or, dès qu’on dispose de lexiques conséquents, on a intérêt à chercher à définir des champs sémantiques constitués de lexèmes ayant entre eux un rapport de recouvrement tel que chacun pourrait remplacer un autre, soit dans un contexte déterminé, soit plus généralement. C’est ce que nous avons fait ici pour quatre racines dont deux, ‘tête’ et ‘oeil’, font partie du vocabulaire de base alors que deux autres, ‘front’ et ‘visage, devant’, en sont généralement exclues. Certaines langues distinguent les quatre lexicalement alors que d’autres en confondent entre eux ou avec d’autres termes encore : c’est ce qui permet de s’apercevoir de l’existence du champ. Ce qui est plus rare et qui intrigue davantage, c’est la présence de racines dont l’étymologie est obscure ayant des sens appartenant à ce champ dans des langues éloignées. Un cas extrême concerne le laal⁴⁴, langue traitée comme un isolat n’appartenant à aucune

⁴⁰ Glottocode leko1246.

⁴¹ Glottocode zand1248, ISO639-3 zne.

⁴² Cf. sama *wūū* ‘corps’ < **yu* à instar du yendang, un remplacement en série ? Mais cf. aussi *jūt* ‘crête (de volaille)’.

⁴³ Le cas du zande est plus compliqué : en plus de son statut de fonctionnelle locative, *tī* est employé comme marque du réfléchi suggérant un passage d’un sens ‘tête’ originel ver ‘personne’, changement qui a pu donné également la racine *zò/ò* ‘personne’ dans une bonne partie des langues oubangiennes à partir d’un emprunt de ‘tête’ en lendu alors que le zande *tīī* ‘tête’ viendrait plutôt du moru-madi.

⁴⁴ Glottocode laal1242, ISO639-3 gdm.

famille, où la racine ‘oeil’ trouve un correspondant très proche en koma-Alantika⁴⁵ avec le sens ‘visage’⁴⁶. Seule une recherche élargie à d’autres champs pourra dire s’il faut reléguer cet exemple au domaine du hasard.

L’identification des constituants des champs sémantiques s’effectue principalement par la découverte de remplacements avérés ou probables mais des tentatives ont été faites d’appliquer ce qui est en quelque sorte la procédure inverse : on cherche des correspondances phonétiques convaincantes et on attribue aux items concernés le statut de champ sémantique, même lorsque le lien sémantique entre eux n’est pas précisable. C’est une procédure qui peut être tentée lorsque la distance entre les langues concernées est considérable et que la proportion du lexique susceptible d’être comparé selon des critères plus stricts de rapprochement sémantique est très basse. Cette méthodologie qui était celle de Bender pour le nilo-saharien (1997) est fondée sur l’idée que les racines présentes à un stade déterminé vont très souvent changer de sens sans être éjectées du lexique. On raisonne alors comme en comparaison multilatérale : rien ne garantit que chaque rapprochement soit valable mais la multiplicité des cas est la preuve qu’un nombre conséquent le sera, ce qu’une recherche approfondie devrait démontrer. L’application de cette procédure en Adamaoua permettrait de constater des rapprochements étonnants comme sama *tùk* ‘oeil’, Adamaoua *s/tu/o(k/i)* ‘oreille’ (où le sama constitue une exception dans une série générale⁴⁷) et burak *lò* ‘tête’, dadiya *lò:l* ‘oreille’ (où les deux langues sont des exceptions dans deux séries générales différentes). La validité des rapprochements de ce genre est bien entendu à débattre.

Cette discussion des innovations sémantiques a été fondée sur le cas canonique de la présence d’un élément insolite dans une grande série comparative mais ce n’est pas la seule configuration possible. Considérons la désignation de la notion ‘eau’ par l’une ou l’autre de deux séries très répandues, **tu* ou **ni* (B&S 1977 : 254). Sans véritables reconstructions, nous ne pouvons savoir ni quelles langues étaient innovatrices ni dans quelle mesure elles l’étaient. On pourrait toutefois envisager pour *ni* un sémantisme ancien comprenant tant ‘eau’ que ‘pluie’ et le remplacement partiel de cette racine par un dérivé d’une racine à consonne alvéolaire sourde initiale + voyelle fermée postérieure signifiant ‘puiser’, donc ‘ce qui est puisé’ pour le sens ‘eau’, créant une opposition lexicale innovante ‘eau / pluie’. Mais on pourrait également supposer que cette opposition était ancienne et que dans une langue descendante déterminée il y ait eu un élargissement

⁴⁵ Glottocode koma1268.

⁴⁶ Dans ce même champ, la racine *laal* ‘tête’ est celle qui apparaît chez ses voisins géographiques kim (kim, glottocode kimm1246, ISO 639-3 *kia* et d’autres non répertoriées par Glottolog). Quant à ‘oreille’, le *laal* a la racine Adamaoua avec la fricative initiale présente en mboumique. On peut se douter que Greenberg aurait rechigné à faire d’une langue le représentant unique d’une nouvelle famille, même si elle impressionne les linguistes de terrain par ses différences d’avec les langues qui l’entourent.

⁴⁷ Le sama a pourtant *táá* ‘oreille’, proche du Tula Wange *kətéélè* (du « Tula-Waja wordlist » de Kleinewillinghöfer 2014) faisant exception dans le groupe tula-waja. On ne peut que signaler en même temps l’exceptionnel *tèkī* ‘tête’ en dama (glottocode dama1267, ISO 639-3 *dmm*) parmi les langues mboumiques occidentales.

du sens de la racine 'pluie' pour englober la notion 'eau'⁴⁸. De plus, rien ne garantit que l'innovation ou la rétention ne se soit pas produite indépendamment dans différents groupes, tout comme le phénomène phonétique de l'avancement de la voyelle fermée postérieure après consonne alvéolaire. L'utilité de l'innovation sémantique pour le classement génétique, constatable dans le cas canonique, disparaît ici.

Au lieu de parler d'« innovations » comme le font B&S, on ferait mieux d'employer le terme « pseudo-innovation » signifiant une conception spéculative au même titre que les « pseudo-reconstructions » qui abondent dans la littérature, telle 'dormir' « in Niger-Congo generally [*sic*] **ĩa* » (B&S 1977 : 252). Les symboles de la formule **ĩa* traduisent le fait que 1) nous constatons dans nos listes une abondance de *la/ra/da* (la voyelle ouverte étant toutefois remplacée parfois par une voyelle antérieure ou postérieure plus fermée) et que 2) la consonne apicale sonore initiale peut parfois être nasale. Cette formule ne représente donc aucunement un lexème qui aurait existé dans une langue ancienne, elle est une simple distillation de traits observés dans les langues actuelles. De la même façon, une pseudo-innovation phonétique n'est que le constat de l'existence dans un nombre restreint de langues actuelles d'une variante apparente d'une racine plus générale à laquelle on attribuera arbitrairement le statut de rétention, toujours sans référence à un hypothétique système phonologique ancien.

5. Du bon emploi des trois méthodologies

Toutes ces considérations expliquent pourquoi la rentabilité de la lexicostatistique à grande échelle paraît si maigre, comme lorsqu'elle nous amène à des conclusions telles que la suivante : « The position of Ijo, like that of Kru, is ambiguous. While its affiliations with S[outh]C[entral] N[iger-]C[ongo] seem reasonably certain, obvious cognates are few. Some evidence links it with East SCNC, other data indicate that it should be classed as a coordinate branch, and a position as an independent branch of Central Niger-Congo or even Niger-Congo is not inconceivable » (B&S 1977 : 251). En effet, lorsque le nombre de termes d'une langue ou d'un groupe de langues pouvant être mis en rapport avec un terme de tout autre groupe reconnu est très petit, la statistique est incapable de trouver une arborescence univoque. Inversement, plus le nombre de termes en rapport est grand, plus une arborescence devient manifeste. C'est pourquoi l'intérêt de la lexicostatistique est restreint aux langues dont la parenté est proche comme nous l'avons déjà indiqué (§3).

L'utilité de la lexicostatistique se limitera donc à la profondeur temporelle atteignable par la linguistique historique dont l'application aux langues africaines, d'abord dans le domaine bantou mais aussi ailleurs, permet de conclure que des états de langue dont l'ancienneté dépassera -4000 à -5000 ans seront difficilement accessibles. En effet, au fur et à mesure qu'on remonte dans le temps, la linguistique historique des langues à tradition exclusivement orale devient moins capable de restituer la matérialité des systèmes phonologiques et de leur diachronie et nous atteignons un stade où nous dépendons entièrement du principe, dont la validité n'est pas garantie, que le passé doit ressembler au présent.

⁴⁸ On remarque dans la liste de mots de base de B&S (1977 : 266), la présence de trois entrées 'année'/'eau'/'pluie', notions qui autorisent souvent des glissements de sens et qu'il faudrait d'une façon ou d'une autre considérer comme un ensemble au même titre que 'abeille'/'miel', vd. B&S 1977 : 242.

Confrontés à cette opacité progressive des systèmes que nous cherchons à connaître, nous avons toujours l'impression d'une étonnante stabilité de certains composants. C'est ce qui permet de réaliser une comparaison multilatérale : ses classements sont fondés sur l'observation d'absences de changement ou de changements familiaux car largement répandus et observables même dans les langues les plus éloignées génétiquement. La comparaison multilatérale est par conséquent la seule méthodologie apte à fournir des hypothèses de classement génétique à grande profondeur temporelle, tout en dessinant les contours des champs où la linguistique historique peut chercher à préciser des faits diachroniques qui nous sont saisissables et qui vont nous aider à mieux comprendre la nature des langues que nous étudions.

Il est évident qu'afin d'étendre le processus de reconstruction en intégrant des groupes qui auparavant auraient semblés trop distants, la linguistique historique aura besoin de vocabulaires et d'inventaires morphologiques de plus en plus riches obtenus moyennant des descriptions approfondies de langues modernes individuelles. Or, tout autant que de l'augmentation du volume de nos connaissances, la reconstruction dans le domaine des langues africaines dépendra de leur qualité, en particulier de l'identification de traits qu'on pourrait qualifier comme étant d'ordre typologique : traits de structure comme l'architecture des systèmes phonologique et tonologique, le sémantisme des sous-systèmes morphologiques, la catégorisation grammaticale des morphèmes et d'autres encore qui peuvent s'avérer être partagés à petite ou à grande échelle. En effet, décrire une langue ne se limite pas à la présentation d'un inventaire de morphèmes avec leur forme phonologique et d'un ensemble de règles permettant de les combiner. Il y a des niveaux plus profonds qui permettent de comprendre le fonctionnement de la langue en synchronie et en diachronie, de dépasser le simple constat et de rendre compte de la façon dont les traits de la langue tiennent ensemble et excluent d'autres possibilités. Selon les langues, certains traits typologiques peuvent perdurer sur de longues périodes ou, d'une façon non encore élucidée, rester latents et réapparaître à une époque postérieure ; d'autres peuvent varier, tout comme les inventaires lexicaux et morphologiques ainsi que la syntaxe. Mais l'application de la méthodologie historique doit restituer une protolangue caractérisable typologiquement et ne doit pas passer sous silence ce genre de faits.

La description linguistique est donc un facteur indispensable à l'amélioration de nos connaissances des langues africaines et les études descriptives approfondies continuent à paraître malgré les nouveaux obstacles à leur réalisation qui surgissent. Or, le nombre de chercheurs qui en profitent pour s'engager sur le parcours de la linguistique historique reste infime. Ce qui était le moteur des études linguistiques jusqu'au début du 20^e siècle semble être progressivement minoré à l'avantage de la typologie, perçue comme plus générale et donnant un meilleur accès à la compréhension de la faculté langagière. En même temps, on peut avoir l'impression que l'horizon ouvert par la comparaison multilatérale est perçu par ceux qui voudraient atteindre l'histoire ancienne des langues comme un raccourci ouvrant une voie vers la formulation de succédanés de postulats historiques qu'on peut améliorer à travers la collecte de toujours davantage d'éléments lexicaux⁴⁹. Si on veut absolument passer avant l'heure à un niveau historique supérieur, on serait

⁴⁹ Dans le cas des langues Niger-Congo au moins, la collecte d'éléments morphologiques tend à être laissée de côté en raison de la complexité des observables.

bien avisé de suivre plutôt la voie suggérée par John Stewart (2002) qui envisageait la reconstruction de protolangues « pilotes » mettant en jeu au moins deux protolangues d'un niveau comparable mais en absence des autres protolangues qui seraient en théorie nécessaires pour atteindre le niveau supérieur visé. En effet, ce procédé serait fondé sur la linguistique comparative et, tout en innovant, resterait dans les voies définies par cette méthodologie.

6. Conséquences générales

A partir de la discussion ci-dessus, nous voyons que chacune des méthodes citées en §1-§5 a sa spécificité, tant par rapport à sa pratique que dans ses objectifs.

- La méthode de comparaison multilatérale permet de dessiner les contours des grandes familles de langues tout en repérant des groupes dont la parenté est si proche qu'on est sûr de pouvoir y appliquer fructueusement la méthode historique.
- Celle-ci, appliquée à un ensemble de langues bien connues et fondée sur des lexiques abondants, accompagnés de préférence par des corpus textuels, nous permet de préciser l'histoire de ces langues : d'abord au niveau des processus de changement qui les ont caractérisées dans le temps, puis éventuellement au sujet de leurs déplacements dans l'espace et de leurs contacts avec d'autres langues. Sous des conditions favorables, notamment lorsqu'une séparation spatiale et culturelle intervient entre certaines langues, elle permettra de constituer des sous-groupes génétiques univoques.
- Si on écarte son emploi dans la glottochronologie, la lexicostatistique ne s'applique fructueusement que dans un domaine restreint, celui des groupes de langues clairement apparentées mais dont l'arborescence établie en fonction des changements constatés est ambiguë⁵⁰. Elle fournit un critère supplémentaire de regroupement, souvent en fonction de facteurs culturels, qui permettra de mieux comprendre les évolutions qui deviendront observables par la suite.

En même temps, on doit reconnaître que chacune de ces trois méthodologies bute contre une difficulté commune : celle des phénomènes de contact de langues. En effet,

- dans le cas de la comparaison multilatérale, comment s'assurer qu'une langue en position de minorité et ayant effectué des emprunts massifs à une langue de prestige, tout en gardant des structures de base présentes avant le contact, n'apparaisse pas à l'observateur comme faisant plutôt partie du groupe auquel appartient la langue dominante ?
- Dans le cas de la linguistique historique, comment intégrer un créole dans une reconstruction classique ? Le problème n'est pas nouveau : l'œuvre de Schuchardt (vd. Schuchardt 1928) soulevait déjà des objections dans ce sens au 19^e siècle.
- En ce qui concerne la lexicostatistique, quel sens donner à une arborescence fondée sur un comptage qui est en principe incapable de distinguer entre les identités dues à l'héritage et celles attribuables au contact ?⁵¹

⁵⁰ Cf. les remarques de Boyeldieu & Cloarec (1986 : 355) concernant l'application de la « dialectométrie » dans le groupe oubanguien.

⁵¹ Vd. toutefois Cloarec-Heiss (2000) pour un procédé statistique plus révélateur intégrant des facteurs autres que lexicaux.

Devant cette situation, on doit se poser la question de l'intérêt réel des arborescences dans le classement général des langues africaines. En effet, s'il s'agit de savoir par où entamer une tentative de reconstruction, ces arborescences ne sont d'aucune utilité car c'est par la simple inspection des données disponibles que nous pouvons déterminer et notre point de départ et les passages successifs à des niveaux plus anciens. On voit difficilement l'intérêt de multiplier et même de généraliser les arborescences que seule une reconstruction pourrait justifier. Greenberg, en tout cas, n'en voyait pas et n'a jamais parlé en termes d'arbres.

Par ailleurs, il semble exister une confusion entre arborescence et histoire. Imaginons par exemple un continuum de cinq parlars différenciés A B C D E disposés sur un axe linéaire dans l'espace et tels que chacun ressemble davantage à celui ou à ceux avec le(s)quel(s) il est en contact direct qu'aux autres. Supposons maintenant qu'à un moment donné C n'est plus transmis. Il restera alors les descendants de deux groupes A - B et D - E. On pourra donc construire une protolange pour chacune, intermédiaire diachronique entre les langues survivantes et leur ancêtre commun. Mais cette arborescence ne représente pas l'histoire, marquée en réalité par la disparition d'un parler, et les protolangues intermédiaires n'ont jamais existé. Elles ne représentent qu'une reconnaissance de l'existence de deux ensembles de traits partagés au moment de l'observation. Le vrai intérêt de la linguistique historique consiste en la formulation d'hypothèses concernant la nature des changements qui ont dû avoir lieu dans les différents domaines - phonétique (neutralisation, assimilation, harmonisation...), morphologique, syntaxique et sémantique - qui expliqueraient les divergences constatées et qui seules représenteraient une réalité. C'est dans ce cadre que nous serons amenés naturellement à situer les changements postulés dans le temps : les anciens qui se sont produits lors d'une communauté précédant la différenciation et qui définissent celle-ci et les plus récents n'affectant qu'une langue ou un sous-groupe reconnu. Mais nous allons nous heurter à tout moment à la difficulté de distinguer les changements induits par contact de langues des changements qui se sont produits indépendamment.

7. Conséquences spécifiques concernant le classement du sama

L'article de Bennett (1983, désormais désigné par B-A), prétend corriger accessoirement un classement effectué par comparaison multilatérale, notamment celui du sama. Son point de départ est une tentative de savoir si les résultats obtenus par G-LA pourraient être atteints autrement :

...is there an Adamawa-Eastern? Do the various groupes of languages assigned to Adamawa-Eastern constitute a viable unit within Niger-Congo? (B-A 24)

Cette notion d'« unité viable » n'est jamais définie en tant que telle, sa définition doit être déduite des procédures employées. La première de celles-ci est une étude lexicostatistique sur le vocabulaire de base des langues Adamaoua-oubanguiennes (= « -Eastern ») identifiées par G-LA permettant de comprendre qu'une unité viable sera celle qui peut être représentée moyennant une arborescence telle que certaines langues apparaissent comme plus proches entre elles qu'avec toute autre langue, chaque regroupement de langues étant une unité viable. Il est évident que les langues Adamaoua-oubanguiennes (dénomination abrégée ci-dessous par Ad-Ou) n'est pas une unité de ce genre car, si elles l'étaient, on s'en serait déjà aperçu et on ne serait pas en train de se poser la question.

Selon les résultats de cette étude (B-A 25-26), il existe un ensemble de langues oubanguiennes et un autre de langues Adamaoua mais plusieurs langues ou groupes compactes de

langues ne peuvent se rattacher davantage de façon chiffrable ni au premier ni au second ni à tout autre regroupement des langues placées dans l'ensemble Niger-Congo : d'un côté le groupe gbaya placé par G-LA parmi les langues oubanguiennes, et de l'autre un groupe toula-longouda, un groupe boua-kulaal et les langues yungur, jen et daka (= sama) parmi les langues Adamaoua (à l'origine, les groupes 1/10, 13, 7, 9, et 3 respectivement chez G-LA).

La démarche suivante de B-A (sans incidence classificatoire) est de recueillir des vocabulaires élargis pour les langues formant l'unité oubanguienne principale, dénommée *kã*⁵², pour déterminer si, comme l'étude lexicostatistique le laissait supposer, il était possible d'y appliquer la méthode comparative. En réalité, seul le groupe que B-A appelle *mundu-gbanziri* (le groupe 5 de G-LA) s'y prête, les correspondances avec les autres groupes étant trop irrégulières ou impossibles à établir par manque de données.

C'est à ce point qu'une dérive s'installe où le sens du terme « reconstruction » est progressivement dilué. D'abord, il est sous-entendu que les termes identifiés comme apparentés dans l'ensemble *kã* entier, dont certains sont reconnus reconstructibles en principe⁵³, sont les représentants d'une seule racine : la notion d'osculation développée par Guthrie (1971) pour le bantou commun est absente. Ensuite, ce procédé est étendu à la totalité des composants de l'Ad-Ou de telle façon que dans tous les cas où le rapport entre des termes ayant le même sens pourrait être interprété comme un apparentement, l'identité historique est supposée et on se permet de tirer des conclusions diachroniques et même de parler de reconstructions en se référant aux termes censés être à l'origine de ces formes diverses.

In the items "breast", "man", and "leaf"...Ubangian shows an initial *k lost elsewhere... The loss of *k seems, in the context of Niger-Congo generally, to be an innovation in the remainder of Adamawa-Eastern, with Ubangian attesting the original form. In the other instances, where there is evidence, Ubangian seems to be the innovator. (B-A 29-30)

C'est ainsi que le constat d'innovations est introduit parallèlement comme garant de la viabilité (terme remplacé par « validité » à partir de B-A 33) d'une unité classificatoire. Ainsi, les observations qu'on vient de détailler suffisent pour affirmer :

On this basis we may be reasonably certain of the status of Ubangian as a viable subgroup of Adamawa-Eastern. (B-A 30)

En effet, si deux groupes de langues A et B partagent une racine avec le même sens et un sous-groupe de A ou de B a une racine différente, on tendra à dire que la racine générale est une rétention et que la racine minoritaire est une innovation même si l'inverse est également possible si l'innovation a eu lieu à l'époque d'une langue commune. Or, de telles innovations sont rares et cotoient de multiples cas d'incohérence dès que plusieurs groupes sont en jeu.

When we turn to Adamawa-Eastern as a whole from these subgroups of Adamawa or Ubangian], we find much the same types of inconsistencies and difficulties. At the higher level, however, it is much less clear that we are dealing with a valid linguistic unit. (B-A 34)

Par conséquent, nous nous retrouvons au point de départ :

⁵² Le nom est tiré d'une reconstruction pour la racine 'sein' dans ce groupe.

⁵³ ...if many of the identifications are dubious, much reconstruction is nevertheless possible. A number of repeated correspondences appear in the data. (B-A 28)

Can, then, the classification Adamawa-Eastern stand as a valid unit? What would be needed to verify the unity of the group, especially given the lack of any lexicostatistical indication of unity, is evidence of internal lexical or phonologic innovation shared by the members of the group, and some indication that no member of the group is more closely related to any other branch of Niger-Congo. (B-A 34)

Au niveau des innovations, il n'y a rien à attendre :

The search for a general Adamawa-Eastern phonologic innovation seems fruitless...

In the area of lexical innovation, I have only...seven items... (B-A 34)

On a alors sept items de vocabulaire dont on n'a pas de nouvelle ailleurs en Niger-Congo mais en absence totale de reconstructions de racines on ne pourra pas garantir que leur présence n'est pas un fait d'ordre aréal. La démonstration de l'unité de l'ensemble va donc dépendre essentiellement du nouveau critère qui vient d'être évoqué : la possibilité d'établir son indépendance des autres grands groupes du Niger-Congo. Or, lorsqu'on regarde le vocabulaire de base manifestement répandu en Ad-Ou, on remarque que de nombreux items trouvent un correspondant en gour et très peu en Benoué-Congo :

...far too large a portion of these characteristically Adamawa-Eastern items are shared with Gur ; while they may strengthen the case of Adamawa-Eastern, they do not argue for its exclusivity. (B-A 35)

Ces rapports sont particulièrement visibles dans un groupe toula-longouda (groupes Adamaoua 1 et 10 de G-LA). Ce qu'il faut alors, c'est un nouvel ensemble de données qui permettra le recours à des calculs lexicostatistiques afin de montrer que le groupe toula-longouda ne se rattache pas de façon prioritaire aux langues Adamaoua. A cet effet on se servira de la trentaine de racines communes au toula-longouda dont on calculera la présence relative en :

- boua/kulaal,
- le groupe Adamaoua principal avec ses deux composants appelés tchamba-namchi et mangbei-mboum⁵⁴,
- le groupe oubanguien entier,
- le gbaya et
- trois sous-groupes gur.

On obtient ainsi - sans réelle surprise étant donné la base de calcul - des chiffres correspondant à « une chaîne dialectale typique » (B-A 36) où le toula-longouda peut être vu comme maillon intermédiaire :

Tula-longouda...is...as close to - or as far from - Gurunsi as it is to Boa-Kula and Chamba-Namshi. (B-A 37)

Cette situation permet d'envisager l'existence un ensemble unique rassemblant le gur et l'Ad-Ou, le Niger-Congo nord central (dont l'acronyme anglais est NCNC) :

if North Central Niger-Congo unity is not proven, it is at least strongly suggested by this study and other data, such as the large proportion of shared Gur-Adamawa-Eastern lexicon already noted. And certainly the exclusivity of Adamawa-Eastern remains very much in question. (B-A 37)

Pour la deuxième fois, la lexicostatistique sert à chiffrer des faits qui avaient déjà été constatés par simple inspection. Or, une fois établi ce rapprochement, on revient encore à l'étude

⁵⁴ Il s'agit des groupes samba dourou et mboumique déjà cités.

lexicostatistique de départ, en insistant sur l'unité du groupe formé par les deux composants principaux de l'ensemble Adamaoua et en voulant montrer par un autre supplément encore virtuel que ceux-ci forment avec le groupe oubanguien une unité supérieure. En effet, si le lien entre les langues Adamaoua était manifeste en termes lexicostatistiques depuis le début de l'enquête, le lien de celles-ci avec les langues oubanguiennes ne l'était pas ; mais

[w]ith the recognition of some of the regular sound shifts^[55] and the addition of new data, however, the link appears closer. (B-A 37).

Les nouvelles données comprennent une liste de neuf termes proposés comme innovations dont seuls deux coïncident avec des membres de la liste de sept fournie auparavant (vd. ci-dessus) pour justifier la validité de l'Ad-Ou entier. Par ailleurs, trois de ces neuf termes ne sont partagés que par l'oubanguien et le mangbei-mboum et un autre, seulement par l'oubanguien et le tchamba-namchi. Or,

[i]t is not clear precisely how the three [groups] are related. They may well constitute a dialect chain. (B-A 37).

En effet, si l'entité désignée NCNC en est une, la nouvelle entité baptisée Cameroun-oubanguien qui en fait partie devrait en être une aussi.

Après ce traitement du plus gros de l'ancien Ad-Ou, il reste à situer les langues restantes mises à l'écart de cette entité centrale par l'enquête de départ, à l'exception du toula-longouda dont le statut intermédiaire entre gur et Adamaoua a déjà été explicité. D'abord le boua-kulaal partage des racines avec le groupe principal Ad-Ou et a été inséré dans la « chaîne dialectale » suggérée par la première enquête supplémentaire :

Cameroun-Ubangian is fairly closely affiliated at least with Boa-Kula. Though from the lexicostatistical viewpoint they are far from close, they do share a number of distinctive innovations^[56]...Statistically, Boa-Kula would appear to be simply a link in the chain connecting Cameroun-Ubangian with Gur. The presence of some of the lexical and phonologic isoglosses noted above, however, suggests rather that Boa-Kula forms part of a larger unit with Cameroun-Ubangian. (B-A 38)

Ces commentaires nous permettent de constater que, selon la méthodologie employée, on peut postuler une « affiliation », en principe génétique, sur la base de la simple inspection d'une liste de vocabulaire de base et sans faire intervenir ni reconstruction ni lexicostatistique, ce qui ressemble fortement à de la comparaison multilatérale parcellaire dépendant d'une application plus générale pré-existante.

En ce qui concerne le falı (groupe dialectal qui constituait le groupe 11 de G-LA), B-A se plaint d'un manque de données mais pressent un lien avec l'Adamaoua central comparable à celui qui a été prévu pour le boua-kulaal :

it could at present be seen as either a branch of Cameroun-Ubangian, perhaps closest to (though seemingly not part of) Chamba-Namshi, or as a branch parallel to Cameroun-Ubangian and Boa-Kula, the similarities with Chamba-Namshi being in part due to geographic proximity.

⁵⁵ Les changements phonétiques en question ne sont pas précisés.

⁵⁶ On en cite onze (B-A 38) dont quatre ont une distribution limitée et quatre autres ont un rapport certain ou possible avec une racine gur.

Passons ensuite aux groupes Adamaoua 7 (yungur, aujourd'hui b̄ana-mboi) et 9 (jen, aujourd'hui le groupe bikwin-jen⁵⁷) de G-LA qui rejoignent le groupe toula-longouda (Adamaoua 1 et 10 de G-LA) en occupant une position intermédiaire entre le Cameroun-oubanguien et le gur. Un nouveau comptage à taille réduite effectué sur une cinquantaine de racines présentes dans au moins deux de ces groupes motive B-A à affirmer que :

I consider these four groups to constitute a unit, which I will call Trans-Benue, consisting of Tula-Longuda, Burak-Jen, and Yungur as subdivisions.

Le traitement du gbaya (groupe oubanguien 1 de G-LA⁵⁸) est à l'opposé de celui accordé aux groupes isolés de l'ensemble Adamaoua.

...most of Gbaya's Adamawa-Eastern vocabulary is shared with Chamba-Namshi and Mangbei-Mbum as often as with Ubangian. (B-A 39)

Cette remarque fait penser d'abord que le statut du gbaya serait comparable à celui du toula-longouda, les deux se trouvant dans une situation de maillon intermédiaire. Or, ce n'est pas le cas car le gbaya, en dehors du nombre réduit de ressemblances lexicales tant du côté Adamaoua que du côté oubanguien, ne présente pas les isomorphismes de morphologie nominale qui relie le toula-longouda tant au gur qu'à d'autres langues Adamaoua. Par conséquent, il conserve son indépendance dans l'arborescence :

...the precise relationship of Gbaya cannot be determined without much further work. That it has developed independently of these three groups^[59] is clear, but until its external affiliations can be clarified it is safest to view it as a branch of North Central Niger-Congo equal to the other three. (B-A 42)

Tous les groupes de l'Ad-Ou sont ainsi casés à la seule exception du...sama dont le statut est vite réglé :

Daka does not seem to be properly grouped with any of the divisions of Niger-Congo here studied...[It] would be better grouped with South Central Niger-Congo. It shows **r̄əm̄* "man", **t̄ie* "head", **ḡæ* "fly", **bara* "two", **t̄w̄ɔ̄na* "five", **ve* "bad", **nyen* "see", and **ba* "come"^[60]; all these contrast with the stems common in the eastern region of North Central Niger-Congo, and certainly with all branches of G-LA's Adamawa-Eastern. They agree instead with stems common in the eastern portions of South Central Niger-Congo. Indeed, the Daka verb *so* "drink"^[61] would appear to be confined to Daka and Greenberg's Plateau languages. Daka should be classified as a member of the Benue-Zambesi subgroup of Eastern South Central Niger-Congo; its precise status within Benue-Zambesi awaits further research.

Précisons d'abord quelques points concernant la liste des racines associées par B-A au BC :

⁵⁷ Glottocode bikw1235.

⁵⁸ Accueilli chez Glottolog sous la forme d'une branche « gbaya-manza-ngbaka », glottocode gbay1279, vd. ci-dessous, §11.

⁵⁹ Le « Trans-Benue », le « Cameroun-Ubangian » regroupant les deux groupes principaux Ad-Ou et le boua-kulaal.

⁶⁰ Respectivement *l̄èrúùm* (peut-être étymologiquement *nè* 'personne + *lúùm* 'mâle'), *t̄ī* (le -e apocryphe, qui crée une fausse ressemblance avec le bantou 'tête', étant une tentative de l'enquêteur historique de représenter la modulation tonale), *ḡèè*, *b̄ààrá*, *t̄úúná*, *v̄èé*, *ɲ̄één*, *b̄áá*.

⁶¹ *Sóó*.

- le sama a un système de numération qui est visiblement d'origine BC⁶² mais qui ne correspond en bloc à aucune langue connue.
- La racine 'venir' est reconstruite par Manessy (1979) pour le gur ; en Oti-Volta (Manessy 1975), on trouve une série partielle **ti/un* avec le sens 'front' que l'on rapprochera facilement du sama 'tête' (cf. ci-dessus, §4).
- Le vocabulaire du sama manifeste, tout comme les autres langues Adamaoua, de multiples correspondances avec le gur (vd. Boyd 1996/97), y compris des traces de classificateurs nominaux suffixés.

Commentons ensuite le principe qui fonde l'argument classificatoire : le vocabulaire de base du sama comprend nettement plus de termes qui ressemblent à des racines caractéristiques du BC que les autres langues Ad-Ou. Aucune méthodologie ne permet d'effectuer une classification génétique sur la base de la simple présence de rapprochements lexicaux entre une langue quelconque et les membres d'un groupe de langues donné. Il faut d'abord prouver que les termes en question appartiennent à un héritage unique et exclusif englobant l'essentiel du vocabulaire, sans quoi rien ne peut exclure la possibilité qu'il s'agisse d'emprunts⁶³.

La sensibilité qui transpire de cet article peut se résumer ainsi : en se familiarisant progressivement avec un ensemble de langues, on se fait une impression globale de leur degré de similarité et/ou de différence. Cette impression est fondée sur du réel, avant tout sur un réel chiffrable. Il faut donc trouver ce qu'on peut compter qui permettra de valider les impressions. Si en fin de compte les chiffres ne sont pas là, on se rabattra sur des innovations lexicales.

Le défaut qui entache le reclassement du sama est donc en fait le non-respect de la priorité que l'auteur donne partout ailleurs à l'approche chiffrée : l'impression d'un parfum BC émanant du sama se comprend mais aucun chiffrage ne vient l'appuyer. Ensuite, si la recherche d'un chiffrage utile s'averait vaine, il faudrait se tourner vers les « innovations » et on découvrirait la présence massive d'« emprunts » au gur en cette fonction (vd. Boyd 1996/1997), dont l'emploi de suffixes de classification nominale. Le reclassement serait alors mis en question. Pour l'instant il reste simplement infondé.

Boyd (1994, 1996/1997) recherche des termes apparentés à des composants du vocabulaire de base du sama dans d'autres langues géographiquement proches ou appartenant à des entités classificatoires dans lesquelles on pourrait envisager de le placer. Boyd (1996/1997 : 30) précise :

If [a wide-ranging cognate] search were to reveal massive cognation and regularity of correspondences in a single language, Chamba Daka would have found a 'nearest neighbor'. If this nearest neighbor

⁶² Sur les dix premiers nombres sont concernés en particulier 'deux', 'cinq' et 'dix'. En effet, 'trois' et 'quatre' semblent présenter plus d'uniformité sur le BC et l'Ad-Ou et 'six' à 'neuf' sont habituellement des composés. Quant à 'un', qui présente une variation extrême, il existe tout de même un rapprochement BC possible (vd. Boyd 1994).

⁶³ Lorsque l'on approfondit ses connaissances des Sama, on apprend que l'existence d'un contact intime avec le BC est cohérente avec leur histoire : ils auraient occupé autrefois un territoire situé à l'ouest de leur emplacement actuel (Fardon 1988) et leur clan royal fait état dans la tradition orale de rapports anciens avec des populations jukun. On pourrait penser à une éventuelle influence de l'« empire de Kororofa » mais cette hypothèse restera vraisemblablement spéculative.

also belonged to some larger linguistic unit, finding it would be equivalent to establishing a continuum between that larger unit and Chamba Daka. As long as such a language cannot be found, there is no such continuum, and Chamba Daka cannot be made a member of any unit.

En ce qui concerne la provenance des différents composants du lexique sama :

Chamba Daka is a highly individuated language with regard to both the proportion of basic vocabulary which remains “untraced” (with no visible cognation) and the processes of phonological change which roots well attested elsewhere appear to have undergone.

Items with Benue-Congo cognation are preponderant but not in the majority, while the proportion of Voltaic items is much higher than would be expected, given the geographical separation and the classificatory positions hitherto suggested.

...it may still reasonably be maintained that the classification of Chamba Daka within Niger-Congo is indeterminate. (Boyd 1996/1997 : 35)

Si nous revenons maintenant à la place de Greenberg devant un vocabulaire de base et que nous jugeons que le sama est bien une langue Niger-Congo, nous nous demanderons, étant donné sa situation géographique, si nous pouvons le reconnaître comme une langue BC et la réponse sera négative. Nous nous demanderons également s’il se rapproche manifestement du groupe identifié comme Ad-Ou et nous verrons que son individualité n’empêche pas son insertion dans un ensemble où règne une grande gamme de variabilité. Le fait morphologique deviendra alors déterminant et nous opterons sans hésitation pour l’Adamaoua. La comparaison multilatérale est ainsi la seule méthodologie capable de classer génétiquement⁶⁴ cette langue. La réalité de son histoire nous restera sans doute inaccessible.

8. La suite d’une sortie de route

Sur la page Wikipedia consacrée au sama en anglais (https://en.wikipedia.org/wiki/Daka_language) on trouve l’affirmation « Bennett (1983) demonstrated to general satisfaction that [Sama] is a Benue-Congo language ». L’emploi de l’expression « à la satisfaction générale » appelle certaines précisions. La proposition de B-A a été reprise par Hedinger (1989) qui y fait référence⁶⁵, ainsi que par Blench & Williamson dans un exposé oral (1987), dans chaque cas sans apport de nouvelles informations. C’est le groupe appelé bantoïde nord⁶⁶ qui s’est révélé particulièrement accueillant pour le sama car il est présent dans son voisinage géographique et en même temps il est particulièrement distendu : Guarisma (1986) établit les taux de ressemblance lexicale entre les langues placées à l’époque dans ce groupe à 80-100/1000 (mambila⁶⁷ et vouté⁶⁸), à 130-180/1000

⁶⁴ Depuis 2024, cette page comporte un lien vers la page de mon propre site qui donne accès à cet exposé.

⁶⁵ « Samba Daka had previously been classified as an Adamawa language but Bennett (1983) pointed out that it is not Adamawa but rather a Benue-Congo language. » (Hedinger 1989 : 424). Le verbe anglais « point out », ‘signaler’, correspond à un fait visible qui aurait pu/dû être remarqué auparavant. Il y aurait donc un cas apparemment unique où Greenberg aurait raté quelque chose.

⁶⁶ Glottocode nort3168.

⁶⁷ Actuellement reconnu sous deux formes, « Western Mambila » au Nigeria (glottocode nige1255, ISO 639-3 mzk) et « Donga Mambila » (glottocode came1252, ISO 639-3 mcu).

⁶⁸ Glottocode vute1244, ISO 639-3 vut.

(mambila et deux parlers kwanja ~ konja⁶⁹) et à 130-200/1000 (vouté et kwanja). Le rajout d'une langue avec un taux comparable n'a donc posé aucun problème.

Pour ma part, lorsque j'ai été sollicité pour fournir le chapitre Ad-Ou (Boyd 1989) pour le même volume synthétique (Bendor-Samuel 1989) où apparaissait l'article de Hedinger et que j'ai présenté ma première version, j'ai été informé par l'éditeur que le reclassement du sama était un fait accompli et que je n'avais pas à en discuter. Pour cette langue, j'ai donc cité les références principales mentionnées ci-dessus et renvoyé au chapitre « Northern Bantoid » (Hedinger 1989) tout en signalant (Boyd 1989 : 182-3) que mes propres vérifications donnaient quelques ressemblances entre le sama et le « nyamnyam » (maintenant nizaa⁷⁰ dont l'appartenance à cet ensemble a été envisagée), et très peu avec le vouté. En prévision du passage référencé ci-dessous, j'ai affirmé dans mon introduction (1989 : 180) « the languages of group 3 (Daka) and 11 (Fali) cannot be attached to any other group in Adamawa, and may be misclassified (see the remarks on Daka below) ».

Après cette publication, j'ai développé avec davantage de précision le raisonnement exposé dans la section précédente et, depuis mes deux publications déjà référencées (Boyd 1994,1996/1997) destinées à montrer l'inutilité de vouloir rattacher le sama à un ensemble particulier autrement que par la comparaison multilatérale, je préface mes travaux sur cette langue avec une déclaration rectificative signalant mon désaccord avec son association préférentielle avec le BC. C'est peut-être pour cela que la page Wikipedia anglaise que je viens de citer déconcerte le lecteur en poursuivant de façon antinomique⁷¹ « though [Sama's] placement within Benue–Congo is disputed ». Toutefois, d'autres sites reconnaissent expressément mon avis : par exemple, « Boyd (2004:195) revises his position on Chamba-Daka and regards Chamba-Daka to be rather a 'peripheral Adamawa language' » (<https://www.blogs.uni-mainz.de/fb07-adamawa/adamawa-languages/>).

Concurremment, Roger Blench a poursuivi le développement de ce thème dans un nouveau sens à la suite de B-A et de Hedinger en plaçant sur son site web personnel (rogerblench.info) un certain nombre d'exposés acceptant comme acquise l'attribution du sama au BC en général et plus précisément au groupe bantoïde nord, en particulier Blench (sans date)⁷², et c'est cette vision qui est acceptée tant par glottolog.org que par ethnologue.com.

Comme la plupart des auteurs de la génération qui a suivi G-LA, Blench ne se propose pas de réaliser un classement *ex nihilo*, autrement dit, de refaire la comparaison multilatérale, mais de toiler le classement proposé par G-LA par des moyens *ad hoc*. Pour cela il adopte une démarche dont on retrouvera une justification dans un article (Blench 2008, apparemment abandonné dans un état incomplet mais cité par la page « Daka_language » mentionnée ci-dessus) consacré à l'ensemble (BC) Plateau auquel il voulait à l'époque associer le sama, placé dans une unité appelée « dakoïde »

⁶⁹ Glottocode kwan1276, ISO 639-3 knp.

⁷⁰ Glottocode suga1248, ISO 639-3 sgi.

⁷¹ Sauf si « placement » dans la citation qui suit est une erreur pour « position ».

⁷² Blench (2014) est plus nuancé et, sans revenir sur le reclassement, le laisse dans le domaine de l'envisageable. On remarque que la référence citée par Glottolog pour la « famille » bantoïde-nord est « Blench, Roger. 2009. The Bantoid Languages. Draft. 88pp. », qui est introuvable. S'agit-il d'un précurseur de Blench (2014) ?

au premier niveau de branchement de l'unité supérieure qu'il appelait « Proto-Central Nigerian » (Blench 2008 : 5).

Il attribue d'abord à G-LA la constitution du groupe dénommé (BC) Plateau avec ses sous-groupes :

the modern subclassification of Plateau derives principally from the work of Joseph Greenberg (1963[1966]) who proposed dividing these languages into seven co-ordinate groups... With numerous emendations and additions these have been reprised in almost all subsequent works... (2008 : 1).

Or, poursuit-il, depuis G-LA les classements sont en mutation constante : elles ne représentent en fait que des impressions du classificateur et ne sont fondées sur aucune « preuve » ou « démonstration »⁷³.

there is no evidence at all that the languages ascribed to Plateau form a distinct group in opposition to Kainji, Jukunoid, Dakoid or Mambiloid (Blench 2008 : 2)

La lexicostatistique prétend au statut de preuve mais, même dans sa version modifiée employée par B&S, ne l'atteint pas en raison de l'insuffisance des matériaux auxquels elle doit s'appliquer :

This paper takes the view rather strongly that lexicostatistics performed on such a problematic dataset are not likely to provide any reliable result (2008 : 2).

Par ailleurs, la linguistique historique qui devrait établir l'unité ou la validité des groupes, c-à-d., montrer que ses membres sont génétiquement plus proches entre eux qu'avec toute autre langue, est impuissante, d'abord parce que les phénomènes de contact de langues obscurcissent les données lexicales :

The whole region is a zone of intense interaction and the following issues must be taken into account:

- a) loans of even fundamental vocabulary are extremely common between adjacent languages
- b) there is no reliable method for recognising such loans

Puis les correspondances morphologiques, en particulier de morphologie nominale, sont incertaines :

- c) loans usually include morphological elements, notably in nominal and verbal plurals
- d) a consequence of this is extreme complexity and diversity in morphology even within one language
- e) this leads to waves of simplification or regularisation of morphology often only partially completed
- f) roots often incorporate fossil morphology
- g) speakers disagree about the 'correct' plural pairing of verbs or nouns
- h) some strategies for regularising morphology spread across regions, rather than being adopted simply within individual languages

Enfin la régularité des correspondances phonétiques est difficile à reconnaître :

- i) sound-correspondences always exhibit numerous 'aberrant' cases

En résumé, le groupe Plateau dans sa totalité est si divers et les descriptions de langues individuelles si peu nombreuses que les données requises pour aborder la construction d'un proto-Plateau correspondant sont inexistantes.

Toutefois, poursuit Blench, tout observateur peut reconnaître dans une multiplicité de langues des racines représentant un même étymon appartenant à une protolangue virtuelle. En explorant

⁷³ Cette observation apparaît inmanquablement chez les opposants à Greenberg, cf. Campbell (1987 : 610) : « Thus [multilateral comparison] is capable of generating vast numbers of intuited similarities... ». Cf. aussi §11 ci-dessous.

l'ensemble des langues contemporaines on obtient ainsi, malgré tout, un accès au lexique de la protolangue, même lorsqu'il n'est pas reconstructible dans la mesure où la « validité » des unités inférieures est indémontrable :

Enough can be hypothesised about common isoglosses to suggest what lexical items can be safely assigned to proto-Plateau, but too little is known about internal subgrouping to put forward regularly reconstructed proto-forms. (Blench 2008 : 3).

Les trois méthodologies habituelles sont ainsi infirmées :

- la comparaison multilatérale ne permet pas de progresser dans le sous-classement, en particulier dans la formation d'arborescences ;
- la linguistique historique ne pourra progresser que lorsqu'on aura écarté tous les écueils quasiment insurmontables posés par les phénomènes de contact de langues (affectant même la morphologie : cf. le point d ci-dessus) et leurs conséquences (concernant la morphologie, cf. le point e) ;
- la lexicostatistique est inefficace en raison de l'insuffisance et de l'hétérogénéité des données lexicales mises à sa disposition.

En revanche, en organisant par ressemblances le lexique commun discerné, on pourra préciser des ensembles de racines caractéristiques qui vaudront classement génétique.

One of the functions of this paper is to present a number of glosses that appear to be characteristic of the group but do not have significant external cognates. Plateau is then assigned a primary branching of Tarokoid; there is considerable evidence for glosses common to the remainder of Plateau, not shared with Tarokoid. The other branches, Beromic, Ninzic, Alumic, Nandic and Jilic, are demonstrably discrete groups...which do not appear to have a hierarchical relationship. (Blench 2008 : 3).

Cette démarche n'est autre que celle de B-A dénuée de tout composant de comptage lexicostatistique. Elle comporte l'identification de racines jugées comparables dans une série de langues ou groupes de langues à rassembler dans une unité supérieure, procédé initial commun aux trois méthodologies courantes. Or, elle ne valide pas en soi l'unité supérieure qui sera postulée en conséquence, elle ne fait qu'afficher une hypothèse à remettre indéfiniment en cause. Ainsi, le toilettage du classement procède par la création d'un flux permanent où chaque lexique nouveau et chaque proposition d'un rapport de parenté donnent lieu à un réaménagement de l'arborescence.

9. Reprendre le chemin

En résumé, nous avons une méthode - la comparaison multilatérale - qui nous permet de distinguer, sur des grands ensembles de langues, celles qui ont un héritage lointain en commun. Cette méthode peut avoir ses limites lorsqu'on repousse très loin les limites temporelles mais elle est efficace au niveau du continent africain.

Si nous nous saisissons d'un groupe de langues très diverses que la comparaison multilatérale a rassemblées et les examinons une par une à la recherche de celles qui se ressemblent le plus, nous allons dénicher sur un extrême celles qui sont des proches parentes et sur l'autre celles qu'on ne peut rapprocher de rien, avec au milieu celles qui laissent apercevoir une parenté difficile à préciser avec l'un ou l'autre des groupes plus compactes. C'est cette procédure qui, d'un côté, permet au comparatiste de définir son domaine et qui, d'un autre côté, motive l'observateur (par

exemple Moñino 1988 et Dimmendaal 2008 pour le cas des langues oubanguiennes) à placer les langues en considération dans des groupes étanches et inaptés à être intégrées dans une diachronie définissable. En ce qui concerne le sama, Boyd (1996/1997) participe de cette dernière tendance, restreinte à l'examen d'une seule langue.

Il ne paraît pas suffisamment reconnu que ces deux approches qui sont la comparaison multilatérale et la linguistique historique représentent deux visions de la réalité et qu'elles ne sont pas en concurrence mais plutôt en complémentarité. Ainsi pour le cas du sama, il nous intéresse de savoir que cette langue appartient à l'ensemble Adamaoua mais que, regardée de près, son affiliation génétique précise est indéterminable. Il ne devrait pas nous surprendre qu'une langue appartienne clairement à une constellation définie et qu'en même temps sa diachronie nous reste obscure et que son étude approfondie ne nous autorise qu'à spéculer concernant les facteurs anciens qui ont pu le conduire à un degré de divergence dont l'interprétation nous restera toujours impénétrable. Pour parler métaphoriquement, le même objet regardé avec un télescope et avec un microscope peut être irreconnaissable.

J'affirme donc que ni les comptages de racines ni l'identification d'innovations hypothétiques, séparés ou ensemble, encore moins les descriptions linguistiques orientées vers la démonstration du caractère unique et incomparable de la langue étudiée, ne peuvent nous apporter un vrai avancement dans le domaine du classement génétique des langues.

Quant à tout essai d'organiser en arborescences un ensemble de langues jugées apparentées par la comparaison multilatérale, il serait préférable qu'on recherche de nouveaux graphismes moins ambigus qui permettraient de mieux visualiser la nature des rapports en jeu et en même temps d'éviter l'association avec une sorte de séparation ancienne des parlars représentés. Les arborescences peuvent exprimer clairement les résultats obtenus lors de calculs lexicostatistiques mais deviennent imprécises et même délusoires ailleurs.

L'emploi d'une représentation par arborescence, adoptée par analogie avec la parenté humaine, a toutefois une longue tradition dans la linguistique historique malgré son implication d'une séparation totale des parlars représentés, rarement réalisée. Connell (2000) a fait une tentative de placement du sama sur une arborescence résultant d'une étude comparative des langues appartenant à l'ensemble mambiloïde où il regroupe essentiellement le mambila, le kwanja et le vouté avec quelques satellites⁷⁴, dans le cadre d'un éventuel élargissement vers un hypothétique ensemble bantoïde nord. Il considère l'appartenance du mambiloïde au bantoïde (glottocode bant1294) comme établie. Il estime (2000 : 211) que « a possible genetic connection between Mambiloid and Dakoid is older than that between Mambiloid and Bantoid...[T]he possibility that Mambiloid is in fact part of South Bantoid needs to be explored...On this view Dakoid would be left as the (as yet unsubstantiated) sole component of Northern Bantoid. The second hypothesis would see Mambiloid...as a branching from a Bantoid node intermediate between Dakoid and South Bantoid. » Autrement dit, sous toutes les hypothèses, le sama restera autant un isolat en BC qu'il l'était en

⁷⁴ Il en exclut de façon provisoire au moins le nizaa et le ndoro (maintenant ndoola, glottocode ndoo1241, ISO 639-3 ndr). On comparera cette distribution avec le mambiloïde de Glottolog (glottocode mamb1309) qui réintègre toutes les langues exclues.

(gur-)Adamaoua. Il n'y a aucun bénéfice à un reclassement en dehors du droit de pouvoir proclamer enfin « Regardez ! Greenberg s'est trompé ! ».

10. Plus = mieux ?

Depuis l'époque de G-LA un grand nombre de langues appartenant à l'ensemble Ad-Ou a fait l'objet d'études descriptives, le sama entre elles. Quel est l'impact de ces études sur le classement effectué au niveau de cet ensemble dans sa totalité ? Ont-elles permis davantage de précision et de raffinement ?

Un premier fait important concernant le groupe Adamaoua est l'existence d'un noyau de langues composé des groupes numérotés 2, 4, 5 et 12 par G-LA. Greenberg ne pouvait pas ignorer la proximité linguistique de ces parlers mais ne commentait pas ce genre de phénomène, évitant ainsi tout appel à la création d'arborescences. C'est seulement à sa suite que B-A a signalé le fait que ces langues qui forment aujourd'hui le groupe « Adamaoua central » offrent un terrain de préférence pour l'application de la méthode historique, d'abord aux langues des groupes 2 et 4 (le samba dourou) auxquelles on adjoindra ensuite celles du groupe 5 (mumuye-yandang⁷⁵). Nous n'avons là aucune avancée véritable par rapport à G-LA.

Le deuxième fait fondamental pour le classement des langues Adamaoua est l'ambiguïté du statut du groupe toula-longouda⁷⁶ : selon Kleinwillinghöfer (2012), les critères qui définissent les langues gur permettent de relier le composant toula-waja⁷⁷ du toula-longouda au « gur central⁷⁸ » et ce rapport est plus net que toute autre affiliation à une langue Adamaoua quelconque. Il écarte ainsi la notion d'une « unité trans-Bénoué » qu'on trouve chez B-A.

On pourrait déduire de ces affirmations que le toula-waja est mal classé et doit être déplacé dans la branche gur ; c'est la solution adoptée par Glottolog. Or, dans ce cas, la notion d'un continuum Gur-Ad-Ou, d'un NCNC, est aussi effacée et les deux groupes reprennent une forme étanche. Nous retrouvons alors l'organisation établie par G-LA avec deux groupes (1 et 10) déplacés d'un ensemble à l'autre.

Un statut de coordonné est donc accordé par Glottolog tant à l'Ad-Ou qu'au gur dans un ensemble supérieur : le « Volta-Congo nord⁷⁹ » (VCN) par la création, au premier niveau de branchement de ce dernier, de deux sous-ensembles, le « Cameroun-oubanguien⁸⁰ » contenant ce à quoi on ne touche pas dans l'Ad-Ou de G-LA, et le « gur⁸¹ » contenant ce à quoi on ne touche pas dans le gur de G-LA, additionné du toula-longouda.

Kleinwillinghofer (2012) ne suit pas cette voie et souhaite continuer à concevoir le toula-waja comme « one of the core groups of a still hypothetical Central Adamawa - Gur ». Ce choix est plutôt judicieux car il existe un fil conducteur, la morphologie nominale, qui relie les ensembles gur

⁷⁵ Glottocode mumu1249.

⁷⁶ Conservé chez Glottolog, glottocode tula1256.

⁷⁷ Glottocode tula1250.

⁷⁸ Glottocode cent2243.

⁷⁹ Il n'y a pas de « Volta-Congo sud », la place étant occupée par le BC.

⁸⁰ Glottocode came1255.

⁸¹ Glottocode gura1261.

et Ad-Ou et fait penser à une explication historique simple, quoique non obligatoire, de la situation observée : des locuteurs de langues ancêtres des langues gur-central actuelles auraient migré vers le sud sur une très longue période pénétrant même la forêt équatoriale. Les langues mba (ou « mbaïques⁸² »), dont le système de classification nominale présente des ressemblances avec celui des langues gur telles qu'il est difficile d'en rendre compte autrement que par une parenté génétique⁸³, seraient donc les témoins de l'aboutissement géographique de ce mouvement migratoire.

En ce qui concerne les différences importantes, tant lexicales que morphologiques, qui distinguent les langues du domaine Ad-Ou, il suffit de regarder la variation qui caractérise déjà les différents dialectes de groupes très compactes comme le toula-waja (Kleinewillinghöfer 2012) pour commencer à prendre la mesure du nombre immense de changements qui a pu se produire pendant cette longue période de migrations successives, la diversification des langues restées dans la terre d'origine étant bien inférieure.

Si nous contrastons cette situation avec celle qui caractérise le domaine bantou où la diversification dans la terre d'origine est bien supérieure à celle observable dans les langues de la diaspora, nous pouvons conclure que, contrairement à la dispersion bantoue, les migrations sur le domaine gur-Ad-Ou n'ont pas pu se dérouler dans un vide linguistique. En effet, des phénomènes de contact de langues ont dû informer en permanence le changement linguistique lors de ces déplacements. Kleinewillinghöfer (2012) parle en effet d'un Sprachbund comprenant des langues tchadiques, nilo-sahariennes et Adamaoua auquel l'ancêtre du toula-waja se serait intégré. On peut aspirer à comprendre dans une certaine mesure le déroulement des influences que ces langues ont exercées et exercent toujours les unes sur les autres car les ancêtres des groupes toula et longouda sont sans doute parmi les derniers à s'être déplacés en direction du sud avant l'arrêt total de ces mouvements de population. En revanche, pour la plupart des autres langues Ad-Ou, les phases de contact depuis leur arrivée en milieu étranger jusqu'aux temps modernes vont rester à jamais inaccessibles.

Quel que soit le cas, nous sommes ici devant le postulat d'un continuum entre deux des groupes majeurs de l'Atlantique-Congo. Or, il faut bien garder en vue la différence entre cette liaison qui estompe une frontière classificatoire apparente et la « liaison » de l'Adamaoua avec l'oubanguien que constitue le gbaya, un isolat dont on cherche les traits qui pourraient bien le rattacher à un côté plutôt qu'à l'autre. Rien ne nous permet de formuler la moindre hypothèse concernant l'histoire du gbaya avant sa diversification mais nous pouvons de manière *a priori* envisager deux possibilités :

- i) la pré-existence en Afrique centrale d'un autre ensemble de langues appartenant à l'Atlantique-Congo ou même à un Niger-Congo plus ancien auquel le gbaya appartiendrait ;
- ii) la génération d'une langue par créolisation dans un processus comparable à celui qui a été proposé pour expliquer l'émergence des langues banda (Cloarec-Heiss 1998) mais sur un temps plus bref et impliquant moins de langues.

En dehors des deux sortes de précision que nous venons d'aborder - signalement de groupes s'offrant à la reconstruction et rapprochement de grands ensembles - qui ouvrent des perspectives de

⁸² Glottocode mbai1246.

⁸³ Je fais part ici de conclusions tirées de mes propres recherches non publiées.

linguistique historique, les tentatives de réaménagement qui ont eu lieu depuis les années 1960 consistent essentiellement à constater que tel parler ou tel groupe de langues est particulièrement original et par conséquent doit trouver sa place ailleurs qu'à côté des autres du même ensemble. Le premier niveau de branchement du VCN chez Glottologue va servir à la représentation de cette démarche en propulsant hors de l'Adamaoua⁸⁴ :

1) le « bua-kim-day⁸⁵ », autrement dit, les anciens groupes Adamaoua 13 et 14 additionnées du day⁸⁶ que G-LA ne connaissait pas. Ces langues très différentes ont comme principal trait commun le fait d'être parlées au Tchad.

2) Le kam⁸⁷ (ancien Adamaoua 8 de G-LA), qualifiée par Lesage (2018) de « unclassified Niger-Congo », aspirant ainsi à une place directement sous Volta-Congo (au dessus de VCN et de BC) qui ne lui a pas été attribuée.

3) Le groupe gbaya (gbaya-manza-ngbaka).

Ces groupes sont ainsi mis en valeur en tant que branches parallèles au Cameroun-oubanguien qui contient la plupart de ce qui reste de l'ancien groupe Adamaoua⁸⁸ : l'Adamaoua central (noyau constitué des anciens groupes 2, 4 et 5), le mboumique ancien groupe 6), le benamboi (ancien groupe 7)⁸⁹, le bikwin-jen (ancien groupe 9) et le kwa (maintenant baa⁹⁰) que G-LA n'intégrait pas.

Cette dissolution de l'Adamaoua constitue une mise nouvelle mise en valeur de ces groupes à ce palier, les transformant en branches parallèles à l'ancien groupe oubanguien entier (moins le groupe gbaya).

Manquent à notre décompte des anciens groupes Adamaoua ceux qui ont été projetés encore plus loin :

1) les groupes 1 (tula-waja) et 10 (longuda, glottocode long1389, ISO 639-3 inu) dont nous venons de parler, rassemblés maintenant en tant que groupe toulal-longouda, l'un des branchements du gur central, lui-même un branchement du gur, branche parallèle au Cameroun-oubanguien ;

⁸⁴ Cette disposition touche également le gur qui se voit retirer l'ancien composant « senoufo » (glottocode senu1239), désormais considéré comme coordonné.

⁸⁵ Glottocode buak1234.

⁸⁶ Glottocode dayy1236, ISO 639-3 dai.

⁸⁷ Glottocode kamm1249, ISO 639-3 kdx.

⁸⁸ Le Cameroun-oubanguien comprend également une branche « Oubangui » (glottocode uban1244) recouvrant la totalité du groupe Eastern de G-LA à l'exception du gbaya-manza-ngbaka, coordonnée avec les langues Adamaoua.

⁸⁹ Notons que Idiatov et van de Velde (2019), poursuivant la méthodologie de Blench (§8 ci-dessus), ont décelé des ressemblances lexicales entre le proto-bantou et le bəna-mboi dans le but d'extraire ce dernier de l'Adamaoua ; bien entendu, pour le repositionner, « [g]eographically, Benue-Congo is the most likely candidate... » (Idiatov & van de Velde 2019 : 10). Pour des raisons inconnues, ils n'ont pas profité de la proposition de Connell (2000) citée ci-dessus, §9, pour commencer à étoffer à nouveau le bantoïde nord en joignant ces langues au sama.

⁹⁰ Glottocode kwaa1262, ISO 639-3 kwb. Il peut s'agir d'une (ancienne) langue de métier (vd. Boyd 1987 : 190).

2) le fali (groupe 11, glottocodes nort2770 et sout2782, ISO 639-3 fal et fil) dont la spécificité, due certainement en grande partie à un contact intensif avec une ou plusieurs langue(s) tchadique(s), est reconnue depuis longue date (vd. Boyd 1989) : il reçoit la mise en valeur maximale en étant placé hors catégorie avec quelques langues ivoiriennes dans « Volta-Congo non classées » (glottocode unun9902) ;

3) et bien entendu le sama (groupe 3) qu'on va retrouver sous « Bénoué-Congo » dans la branche bantoïde nord du bantoïde.

On voit dans cette restructuration trois processus :

- l'éparpillement par déconnexion des langues environnantes (souvent dans une mouvance « avec des données insuffisantes, Greenberg ne pouvait que se tromper ») ;
- la mise en valeur par transformation en branche parallèle ;
- le rabaissement automatique de tout ce qui ne subit aucun processus de mise en valeur, particulièrement remarquable dans le cas de l'Oubangui dont le contenu aurait pu être rehaussé d'un palier en tant que trois nouvelles branches du Cameroun-oubanguien : le zandic⁹¹, le ngbandi-mongoba-kazibati⁹² et, disons, un Oubangui central avec deux branches : l'un regroupant sereïque, mundu-baka et mbaique⁹³, l'autre réservée au bandaique⁹⁴.

Ce relevé nous montre que, lorsqu'on intervient sur une comparaison multilatérale, on agit soit dans le sens du regroupement de langues qui incitent à entreprendre une étude comparative, soit dans le sens de l'individualisation de (groupes de) langues qui ne se prêtent aucunement à l'approche comparative avec nos connaissances actuelles. Dans le premier cas, on crée une arborescence en fonction d'une échelle de proximité qui définit la séquence des protolangues à reconstruire. Dans le second, on peut vouloir projeter les langues récalcitrantes sur cette même arborescence à un niveau supérieur ou les expulser tout simplement du groupe d'origine.

Le problème de la représentation par arborescence est qu'elle est contrainte par cette dualité qui impose une césure dans la méthode historique. Ainsi, si nous visons la reconstruction d'une protolangue « samba dourou », nous voudrions effectivement confronter celle-ci, au niveau suivant, à une protolangue mumuye-yandang que nous aurons reconstruite par ailleurs. Mais une fois que nous aurons notre proto-Adamaoua central, nous n'avons aucune idée des (proto)langues à adjoindre parmi les groupes placés par le classificateur du moment sur le même niveau de branchement afin de poursuivre notre régression dans le temps.

Une deuxième complication est liée aux phénomènes de contact de langues : si un créole devient autonome ou une langue véhiculaire devient première langue transmissible, une arborescence les représentera sous forme de branches parallèles et sera par conséquent ambivalente, donc inadéquate. S'il est impossible de dire à quel niveau chaque composant va rejoindre les autres, la protolangue n'est qu'une abstraction théorique impossible à rattacher aux langues réelles. Celles-ci de leur côté deviennent un simple assortiment qu'on peut rassembler différemment selon les critères

⁹¹ Glottocode zand1246.

⁹² Glottocode ngba1290.

⁹³ Glottocodes sere1262, ngba1292 et mbai1246 respectivement.

⁹⁴ Glottocode band1341.

appliqués, soit linguistiques (structure syllabique, complexité morphologique, organisation sémantique...), soit sociolinguistiques (diversification, rapport à un status social, mixage...).

Devant cette situation, il est difficile de vouloir maintenir la représentation par arborescence autrement que comme une sorte de métaphore : « à quelque niveau que ce soit, nous ne savons pas grand-chose de la diachronie des langues que nous classons, alors ayons recours à l'arborescence comme expression d'un rapport approximatif indéfiniment modulable en attendant des informations plus fiables ». Comme nous l'avons déjà évoqué au §3, un besoin existe de recherches orientées vers le développement d'un modèle topologique plus satisfaisant, ne fût-ce qu'une représentation tridimensionnelle lisible et informative situant les différents éléments à des distances les uns des autres, modulables en fonction d'un ou de plusieurs critères précis.

On souhaite bien entendu une bonne vision d'ensemble des rapports tant de la parenté des langues africaines que de ce relevant de la notion essentielle de Sprachbund ; mais la seule façon de faire progresser nos connaissances de la diachronie de ses langues passe par des descriptions de parlars contemporains auxquelles on appliquera ensuite la méthodologie comparative en l'adaptant à nos possibilités et à nos besoins (vd. encore Stewart 2002).

11. Plus sur le sama en particulier

Je résume les conclusions formulées ci-dessus concernant le sama.

La simple inspection de son vocabulaire de base montre que cette langue n'a aucun parent proche parmi les langues avec lesquelles la comparaison multilatérale permet de la grouper (le groupe Adamaoua de G-LA). En cas de projection des résultats de cette comparaison sur une arborescence, la position du sama sera indéterminée.

Le lexique du sama comprend un certain nombre de termes (sans doute plus que dans les autres langues du même ensemble) repérables comme caractéristiques des langues BC de sa région (cf. ci-dessus §7). Le peu qu'on connaît de l'histoire ancienne des Sama est compatible avec l'hypothèse d'une domination politique exercée par un peuple parlant l'ancêtre d'une ou de plusieurs de ces langues (pendant l'empire de Kororofa⁹⁵ ?) qui pourrait expliquer la présence d'éléments lexicaux d'emprunt. Cette composante lexicale a pourtant été interprétée non pas comme un emprunt mais comme la preuve d'un héritage, même si cette langue ne paraît pas plus proche de celles de son point de chute en BC qu'elle ne l'est de celles de son groupe d'origine.

Or, l'absence totale, non seulement de tout préfixe de classe nominale mais de tout signe qu'il y ait eu, sous quelque forme que ce soit, un seul préfixe de classe n'est pas favorable à cette option car, à l'opposé du BC, c'est en NCNC que la suffixation est préférentielle et les cinq pluriels nominaux supplétifs attestés en sama comportent bien des vestiges de suffixes de classe anciens.

11.1. Morphologie du système verbal

Signalons maintenant un autre trait morphologique du sama beaucoup plus difficilement décelable au niveau de la simple inspection du vocabulaire de base que les traces de suffixes de classes nominales : la présence d'un paradigme de suffixes de dérivation verbale, rudimentaire mais très productif, qui rappelle de manière frappante des systèmes bien représentés dans les langues BC.

⁹⁵ Cf. <https://en.wikipedia.org/wiki/Kwararafa>.

En particulier, ce paradigme, ainsi que le paradigme de marques TAM du prédicat, ressemble étonnamment à celui du Bafia (langue bantoue A.50, glottocode bafi1243, ISO 637-3 ksf). J'ai examiné ce cas en détail dans un article non publié car rejeté par tous les référents. Ces derniers partageaient en effet du positionnement théorique suivant :

1) on peut affirmer qu'un trait *t* puisse être reconstruit pour la protolangue *P* d'un ensemble *L* de langues et que *t* soit présent également dans une langue *x* ne faisant pas partie de *L*. Mais si *x* et *L* font toutefois partie d'un ensemble ayant comme protolangue *Q*, bien plus ancienne que *P*, alors la présence du trait *t*, tant dans *x* que dans *P*, relève d'un héritage.

2) En supposant que le hasard soit exclu, si *t* n'est pas un héritage, il doit être un emprunt. Or, tout emprunt est localisable dans l'espace/temps et sans cette localisation, aussi improbable que l'hypothèse de l'héritage puisse paraître, elle s'impose.

3) Par conséquent le paradigme de dérivation verbale du sama est un héritage et le sama est une langue BC.

Il est vrai que ce phénomène mérite une explication mais le recours à un simple reclassement du sama dans la branche BC n'est pas plus obligatoire ici que dans le cas du zandé qui présente, à la différence des autres langues oubanguiennes, un système comparable (vd. Boyd 1995). Les deux langues, sama et zande, se ressemblent d'ailleurs par le fait d'avoir greffé ce système sur un système pré-existant devenu improductif. Il peut donc s'agir de la ressuscitation d'un procédé ancien de dérivation moyennant l'incorporation d'une nouvelle morphologie productive, peut-être au travers d'un contact avec une autre langue dont la descendance actuelle n'est toujours pas identifiée.

11.2. La question du renouvellement de la morphologie nominale au niveau régional

Il n'est d'ailleurs pas exclu que, dans des langues bantoïdes nord, il y ait eu génération d'un ensemble de préfixes de classification nominale par voie de contact. Les systèmes actuels ne seraient donc pas composés des fossilisations de morphèmes autrefois productifs et généraux mais plutôt d'un ensemble de marques nouvelles importées afin d'effectuer un classement sémantique⁹⁶. Cette hypothèse cadrerait bien avec 1) le nombre réduit de préfixes attestés dans chaque langue ; 2) l'absence d'appariements singuliers/pluriels, inutiles pour la simple classification sémantique ; 3) la généralité de ce mode de classification, alors que, dans les langues avec des systèmes hérités, les critères de classification sémantique sont flous en dehors des noms d'être animé et de masse ; 4) la difficulté à établir des correspondances phonétiques entre les préfixes associés dans chaque langue à des radicaux lexicaux comparables, avec la possible exception de *mV-* pour les noms de masse.

On remarque d'ailleurs la présence, notamment en vouté mais peut-être aussi dans d'autres langues bantoïdes nord, de vestiges de marques de classe suffixées qui représenteraient alors un système plus ancien sur lequel toute préfixation serait une greffe. Or, cette question de la génération de marques de classification nominale non héritées est un thème rarement abordé dans le domaine de la linguistique africaine, sans doute principalement parce qu'il est difficile, dans un contexte de transmission uniquement orale, d'apporter des preuves d'une hypothèse quelconque. D'autres cas envisageables dans ce cadre ne manquent pourtant pas en Afrique centrale : on peut signaler par

⁹⁶ A ce sujet, vd. l'omniprésence des critères sémantiques chez Martin (sans date).

exemple l'existence d'un petit paradigme de suffixes nominaux de la forme $-CV^{97}$, sans opposition singulier/pluriel, dans le mboumique, visible en particulier dans les nominaux de la forme CVCCV. Ce paradigme comprend un suffixe $-mV$ correspondant aux noms de masse. Greenberg (G-LA 11) attribue les C_2 de certaines racines CVC à un ancien système suffixal disparu dans les langues modernes de ce groupe (cf. note 19). Les suffixes $-CV$ sont-ils aussi résiduels ou témoignent-ils plutôt de la régénération minimaliste d'un nouveau système ? Dans ce dernier cas, le système est-il d'origine strictement interne ou s'agit-il d'un phénomène de contact avec une autre langue toujours indéterminée ?

De même, on notera la présence d'une classification nominale en tiba (*gàà* pour ses locuteurs, glottocode gaaa1245, ISO 639-3 ttb), langue Adamaoua central⁹⁸ dont le classement généalogique est évoquée ci-dessous, §12) moyennant l'emploi de deux préfixes, *á(à)*- 'être animé', *à*- 'être inanimé', qui ne sont pas intégrés à la racine, étant même facultatifs dans plusieurs contextes morphosyntaxiques⁹⁹. Ce phénomène, ajouté aux autres cités ci-dessus, nous amène en tout cas à nous interroger sur une éventuelle facilité des langues de la région, en particulier celles qui ont dû perdre un système suffixal antérieur¹⁰⁰, à relancer une classification nominale par préfixation. Si cette hypothèse devait être retenue, un trait qui unirait les langues bantoïdes nord - ou au moins certaines d'entre elles - serait le fait d'avoir été rapprochées de l'ensemble BC par « mixage » plutôt que par héritage.

On voit à travers ces quelques remarques que les connaissances approfondies acquises par la description des langues « périphériques » semblent sans impact sur leur classement par comparaison multilatérale mais aiguës notre conscience de phénomènes attribuables au contact de langues. Devant une langue quelconque, nous pouvons nous demander si nous observons soit la présence de traces de contact suggestives d'une transformation moyennant une période prolongée de bi- ou de multilinguisme, soit des signes de génération par un processus de créolisation, c-à-d. une langue seconde remodelée devenant langue première (vd. pour le bandaique, Cloarec-Heiss 2000). Il n'est pas sûr qu'à une certaine profondeur temporelle, nous puissions distinguer les deux cas. On peut même envisager une zone d'ombre où, sous les effets du contact, une langue atteindrait un point de

⁹⁷ C = consonne, V = voyelle.

⁹⁸ « Dakaoïde » (dako1256) pour Glottolog suivant Blench (sans date1).

⁹⁹ D'après Güldemann (2021) la classification nominale sur la base du trait \pm animé serait localisée principalement en Afrique centrale. Le groupe linguistique le plus proche géographiquement utilisant cette forme de classification est celui des langues gbaya qui emploient la pronominalisation à cet effet. La présence d'une classification par préfixes en tiba où elle est en principe d'origine récente, étonne et reste pour l'instant inexplicée.

¹⁰⁰ Le cas de perte d'une flexion suffixale semble avéré en Adamaoua dans le groupe samba dourou. En effet, les langues de type koma (Dieu & Perrois 2016), répertorié par Glottolog sous le nom gimme (glottocode gimm1241, ISO 639-3 kmp), possèdent un système de ce genre, présumé hérité, qui est absent dans les autres langues du groupe. Celles-ci n'ont pas adopté pour autant un système préfixal innovant, cf. le cas du dii (Bohnhoff 2019, glottocode diii1241, ISO 639-3 dur) dont le comportement calque celui du sama : absence de préfixes de classification et emploi d'un pluriel universel suffixé $-v\#$, identifiable en principe avec le pluriel sama $-b\bar{u}$.

basculement, les changements opérés devenant si importants que la langue défavorisée apparaît désormais comme appartenant au même groupe génétique que la langue dominante. Actuellement nous ne pouvons pas affirmer avec certitude qu'un tel cas se soit produit mais nous ne pouvons pas l'exclure.

12. Post-scriptum pré-polémique

12.1. Dire ce qu'on voit et expliciter comment on fait pour le voir

Au cours de cet exposé, j'ai soutenu que le reclassement du sama relève d'un rejet de la comparaison multilatérale suivi de tentatives de récupérer - moyennant des raccourcis méthodologiques - plus de connaissances que cette méthodologie ne peut elle-même fournir, tout en conservant ses acquis en arrière-plan. Il convient de jeter un regard maintenant sur la raison principale pour laquelle la gestion de la comparaison multilatérale inspire ce rejet et d'envisager les moyens de pallier cette réaction.

Ci-dessus, §7, j'ai employé le mot « impressions » pour parler des perceptions dont l'enquêteur se méfie, puis revient dessus en doutant et en voulant une confirmation par le chiffrage. Face à cette évaluation négative des impressions, Newman (1995) utilise le mot « feel » associé à la notion de « insight » dans un sens positif pour appréhender la méthodologie de Greenberg. Le lecteur pourrait s'insurger : c'est comme si on voulait faire de l'intuition le moteur d'une procédure scientifique et instaurer en conséquence une autorité incontestable émanant d'individus surdoués (cf. les remarques sur la notion de « preuve » ci-dessus, §8).

Pour ma part, je prends plutôt le parti de Newman et je cite ici anecdotiquement quelques expériences personnelles qui montrent comment les impressions peuvent être mis à profit si l'observateur veut bien abandonner sa méfiance.

- J'ai participé à la phase initiale d'un projet, Corpafroas (<http://corpafroas.tge-adonis.fr/>), consacré à la famille afroasiatique car, tout en étant spécialiste des langues Ad-Ou, j'ai travaillé pendant plusieurs années sur la langue bata (glottocode bata1314, ISO 639-3 bta) qui appartient à l'ensemble biu-mandara sud (glottocode sout3145) des langues tchadiques. J'ai été frappé, lors des échanges avec mes collègues travaillant dans d'autres branches des langues afroasiatiques (berbère, sémitique, couchitique...), par les ressemblances manifestes, en particulier au niveau de l'architecture des systèmes phonologiques, de la phonotactique et de la morphologie, qui distinguent ces langues de celles de mon domaine principal de recherche. Pour continuer dans la métaphore non scientifique, c'était comme si j'explorais d'autres pièces d'une seule maison. Je ne pouvais pas douter chez qui j'étais.
- J'ai eu l'occasion de travailler de façon intensive avec la langue sèmè du Burkina Faso, appelée siamou (glottocode siam1242, ISO 639-3 sif) en dioula, la langue véhiculaire régionale. Person (1968) la reconnaissait comme une langue kru en le contrastant avec des descriptions d'autres langues de ce groupe. Aujourd'hui un regard sur les systèmes tonals du sèmè et des langues kru, en particulier des parlers du nord (cf. Bearth & Link 1980), ne laisse pas douter qu'il s'agit de parents¹⁰¹. Pourtant,

¹⁰¹ Un contravis : on sait que les Seme ont vécu assez longtemps parmi des peuples mandé et que leur langue en montre les empreintes. Ayant très peu de connaissances en langues mandé, j'ai demandé à Denis Creissels si

Güldemann (2019) reconnaît les preuves de l'appartenance des langues kru à l'ensemble Niger-Congo comme « prometteuses » à l'exception du « siamou »¹⁰² et son avis est suivi par Glottolog qui l'exclut du Niger(~Atlantique)-Congo.

- Si on connaît quelque peu les langues mumuye (« mumuyique » pour Glottolog, glottocode mumu1250, cf. Shimizu 1979) et qu'on enquête sur le tiba/*gaa* (cf. ci-dessus §11), je ne vois pas comment on peut hésiter à placer cette langue dans le même ensemble malgré la présence d'emprunts lexicaux massifs au sama au milieu duquel elle est enclavée.

- Et enfin, comme nous l'avons déjà indiqué ci-dessus, §8, si on a une certaine familiarité, bien au delà de celle qui suffit au comparatiste multilatéral, avec les langues bantoïdes qui n'appartiennent pas au « bantou étroit¹⁰³ » ainsi qu'avec la grande variété des langues Adamaoua et qu'on est confronté à la question : dans quel ensemble faut-il placer le sama ? on n'hésitera pas à choisir l'Adamaoua.

Ce sont ces réactions qui m'incitent d'emblée à reconnaître une faculté de vision que certains pourraient qualifier de non scientifique. Il faut par conséquent qu'elle fasse l'objet d'une « mise entre parenthèses » (cf. Husserl 1986) pour qu'on en fasse une analyse raisonnée.

Pour ma part, je prétends la concevoir comme une faculté de saisie de configurations définissables et contrastables du réel. Elle doit donc s'exercer sur des réalités qu'on peut désigner et expliciter. Il faut ensuite que la justesse de la saisie - ou plutôt de sa verbalisation - puisse être évaluée : l'autorité ne peut pas être arbitraire. Et bien entendu, il sera toujours possible de formuler le reproche que les configurations saisies sont un appauvrissement de la réalité, qu'en donnant une prise sur la réalité elles dissimulent sa richesse.

12.2. Ce que Greenberg a vu et comment il l'a vu

Avec une définition si large, cette faculté a nécessairement une gamme d'avatars dont je ne citerai que deux :

1) La première nous place dans le cadre de la structuration de la pensée lors de théorisations attribuables à un individu. L'association d'un nom avec une découverte scientifique est relativement fréquente, telle Newton et la loi de la gravitation ou Heisenberg et le principe de l'indétermination. Ces « découvertes » n'apparaissent pourtant pas *ex nihilo*, elles consistent en une formulation cohérente intégrant des idées et des observations contemporaines partagées.

Dans ces cas relevant du domaine de la physique, les nouvelles formulations peuvent comporter des aspects contre-intuitifs, en désaccord avec l'univers rationnel et perceptuel de l'être humain (force sans support, réel imprécisable...) et c'est alors l'affranchissement de ces carcans intellectuels qui permet l'avancée scientifique : les énoncés théoriques émis appellent des réflexions ultérieures et des vérifications qui vont les préciser et les développer.

la phonologie du sèmè, caractérisée par divers phénomènes d'harmonie phonotactique, correspondait à un système mandé représentatif. Il a répondu par une question, « c'est quoi un système mandé ? » comme s'il niait que dans cette famille on puisse avoir la sorte de sentiment dont je fais état.

¹⁰² Vd. aussi Vogler (2019)

¹⁰³ Glottocode narr1281.

Dans d'autres domaines, la découverte va à l'encontre d'une idéologie pré-existante dont elle révèle les contradictions. L'innovateur se libère et se détache d'une pensée explicative précédente en fonction de nouveaux constats et de nouveaux courants d'idées, ainsi Darwin lorsqu'il conçoit l'évolution des espèces en fonction de la sélection naturelle.

Dans les mathématiques également, on voit chez certains individus une coalescence d'une gamme de réflexions qui ouvre une voie auparavant inaperçue qui pourra être exploitée collectivement par la suite : pensons par exemple à Thom et la théorie des catastrophes.

Il est presque universellement accepté que, dans tous ces cas et dans d'autres encore, on puisse parler de la « vision » ou du « pouvoir de synthèse » de celui qui a été le premier à formuler l'expression théorique correspondante. L'application de ces dénominations à la démarche de Greenberg dans le classement génétique des langues paraît raisonnable dans la mesure où, en évacuant des préjugés déformants (cf. Newman 1995), elle a donné d'abord un nouveau cadre pour la considération et l'exploration des langues africaines dans leur totalité.

2) Plaçons-nous maintenant sur un tout autre plan : celui de l'adaptation de la pensée à une réalité nouvelle. Peu sont les professeurs de langues étrangères qui refuseraient de reconnaître qu'étant donné un échantillon arbitraire d'élèves âgés de 14 ans ou plus, il y aura parmi eux une échelle d'aptitude à l'apprentissage qu'aucune technique pédagogique ne pourra éliminer. Ce sont des faits qui relèvent plutôt de la compétence linguistique individuelle, d'une capacité relative d'assimilation.

Il est peut-être moins évident mais toujours envisageable de ranger dans cette catégorie la facilité avec laquelle Greenberg localise des signifiants apparentés dans la masse d'informations consultées. En effet, nous apercevons à la lecture de G-LA un ensemble de lexèmes, de morphèmes et de paradigmes qui servent à étayer son classement mais nous n'avons pas accès au processus de découverte de traits structurels et de types de correspondances caractéristiques des différentes familles qui a conduit au choix des données incluses dans la présentation finale. Il est d'ailleurs possible que Greenberg même n'ait jamais verbalisé ce processus qui se serait alors déroulé de façon intuitive. De telles circonstances pourraient nous autoriser à lui attribuer une compétence particulière dans cette activité que d'autres n'auraient pas su exercer.

Or, qu'on regarde les contributions de Greenberg en tant que production de discours structurants ou en tant que capacité à appréhender des configurations pertinentes, il faudra toujours mettre à l'épreuve la réussite de ces procédés cognitifs. A cet effet, ses critères et leur application méritent bien d'être décelés, décortiqués et débattus. Ce processus d'analyse et de vérification n'a pourtant pas eu lieu comme il aurait dû à cause de la polarisation que la comparaison multilatérale a provoquée : soit on accepte sa validité et on y fait référence sans autre considération, soit on lui refuse toute validité : « seule la linguistique historique est sérieuse ». La solution hybride qui consiste à la valider de fait, tout en essayant de la mettre à mal en trouvant des « erreurs » de classement ou des « améliorations » moyennant le réaménagement des groupes ou l'injection de sous-classements, n'éclaire pas davantage son essence.

12.3. Ce que je vois du sama et de ses voisins et comment je vois leur classement

Revenons maintenant au tout premier niveau d'appréhension et considérons les conséquences pour notre propos de ce petit précis de philosophie de la science, à savoir, pour la détermination de la situation classificatoire du sama et de certains de ses voisins. Lorsque je dis, par exemple, en §12.1 que telle langue (en l'occurrence le tiba et vraisemblablement aussi le dong, glottocode dong1293, ISO 639-3 doh) est à ranger aux côtés de telles autres (en l'occurrence des langues mumuye-yandang) plutôt qu'avec le sama, bien singularisé par rapport à celles-ci, il est évident que je le fais en revendiquant une certaine compétence qui pourrait relever de l'ordre de la « vision ». En effet, mon choix peut avoir une qualité intuitive ; il est néanmoins fondé sur des faits qui doivent être précisés en cas de contestation. Il reviendra bien entendu à ceux qui me succéderont et qui regarderont de plus près ces langues de dire si les ressemblances qui fondent mon avis sont décisives et justifient mon évaluation.

Pour le moment, contentons-nous de contraster les configurations du classement du tiba et du dong d'un côté et du sama de l'autre. Si nous restons dans le cadre lexical, il y a un accord général sur le fait que chacune de ces trois langues a des éléments de vocabulaire qu'on va rapprocher d'éléments présents dans l'un ou l'autre de deux ensembles de langues bien distincts. Le fait du classement consiste à dire en gros que les ressemblances avec l'un de ces deux ensembles sont dues à un héritage alors que celles qui rapprochent la langue en considération avec l'autre ensemble résultent du simple contact. Dans le cas du sama, les deux ensembles sont d'un côté les langues BC, les plus proches géographiquement étant les sous-ensembles

1) bantoïde dont le bantoïde nord et, dans le bantoïde sud, les langues bantoues A et les langues des Grassfields¹⁰⁴,

2) (BC) Plateau et

3) jukunoïde

et de l'autre côté les langues de l'ensemble Cameroun-oubanguien dont celles de l'Adamaoua central sont les plus proches.

Dans le cas du tiba ou du dong, il s'agit des langues Adamaoua central d'un côté et de l'autre, de l'ensemble ayant comme élément unique le sama.

Une différence majeure sépare donc ces deux cas : dans le cas du tiba ou du dong, les deux ensembles de référence sont de petite taille et pratiquement tous les composants du vocabulaire de base y trouvent un correspondant manifeste. Dans le cas du sama, en revanche, les ensembles de référence sont énormes, les correspondances s'éparpillent entre différents sous-ensembles et il reste une partie du lexique de base qu'il est difficile de situer et qui n'a aucune « source » claire, qu'elle soit d'héritage ou de contact (cf. §7 ci-dessus).

De même, dans un cas comme dans l'autre, nous supposons que la configuration observée témoigne de l'existence actuelle ou passée d'un bilinguisme et qu'en conséquence la plupart des éléments de vocabulaire indicatifs de contact proviennent, à proprement parler, non pas d'un

¹⁰⁴ Cf. chez Glottolog les ensembles bantou A-B10-B20-B30, glottocode bant1295, et Grassfields étroit, glottocode narr1282.

« ensemble » de langues mais d'une langue précise, potentiellement l'ancêtre d'une langue existant aujourd'hui.

Ensuite, les deux cas diffèrent à nouveau par le fait que, dans le cas du tiba ou du dong, nous pouvons identifier la source de ces éléments de contact : il s'agit du sama. En revanche, le stock lexical du sama qui rappelle le BC n'est attribuable à aucune langue ou groupe de langues actuelles précis.

De la même façon, en ce qui concerne le lexique hérité du tiba et du dong, nous pouvons le mettre en rapport avec ce que nous pouvons reconstruire à partir d'un ensemble diversifié mais cohérent sous forme d'un mumuye commun ou, étant donné l'existence en dong d'un petit ensemble de termes partagés avec le yandang, d'un mumuye-yandang commun. Mais le stock lexical du sama qui est apparemment d'origine VCN ne peut pas être retrouvé dans une protolangue unique de proximité. On le retrouvera éparpillé autant dans les langues Adamaoua central que dans le mboumique et ailleurs en Cameroun-oubanguien, une partie étant localisable même dans le proto-otivolta et le proto-gur de Manessy (1975, 1979), cf. Boyd 1996/97.

Le seul cas de partage lexical conséquent entre le sama et une langue de proximité concerne le samba-leko. Le stock commun est en effet important - pourvu qu'on sorte du vocabulaire de base - mais est presque entièrement absent des autres langues apparentées de près au samba-leko, ce qui laisse penser qu'il s'agit principalement d'emprunts effectués par celui-ci au sama. En effet, si à l'intérieur de l'ensemble ethnique sama/samba, ce sont les Samba qui exerçaient une prédominance rituelle, ce sont bien les Sama qui détenaient le pouvoir politique associé à une supériorité numérique (vd. encore Fardon 1988).

On pourrait envisager d'appliquer le même paradigme au tiba et au dong mais leur situation linguistique diffère par rapport à celle du samba-leko dans la mesure où le composant sama de leur lexique atteint le vocabulaire de base. Si ces deux langues sont, comme il paraît, associées au métier de forgeron, ce seraient des populations spécialisées dans cette fonction qui auraient été absorbées par les Sama et isolées de leur ethnie d'origine. Mais alors les Dong, pour une raison indéterminée, seraient revenus ensuite parmi des populations proches de cette dernière.

Cette exposition focalisée sur le lexique doit évidemment être élargie afin de prendre en compte des faits morphologiques et syntaxiques. Or, si j'ai examiné divers phénomènes observés dans ces domaines en sama dans des publications que je ne dénombre pas ici, je n'ai publié qu'une esquisse du tiba (1999) et n'ai jamais enquêté sur le dong : je ne le connais que par le lexique de Blench (2006). Mon exposé sur la question s'arrête donc à ce stade et devra être confirmé ou infirmé à l'avenir.

13. Postface polémique

13.1. Le filtrage

En §12 j'ai voulu faire comprendre que cet exposé prend place dans une tradition de réactions conflictuelles à la comparaison multilatérale. Il a par conséquent une facette qu'on peut qualifier de psychosociologique : je prétends concrétiser un avis dans un discours raisonné et cohérent dans le but de convaincre mes pairs qui n'ont pas encore pris le parti de rejoindre mon point de vue. Cette intervention a lieu dans un contexte particulier : en effet, en me focalisant sur le

thème du classement génétique du sama, je ne me limite pas à douter de la validité d'une seule étude mais de celle de toute une bibliographie qui se réfère à cette étude, bibliographie dans laquelle je me trouve moi-même inclus faute d'avoir su m'opposer à la thèse qu'elle véhicule au moment opportun. De plus, le reclassement proposé et la bibliographie afférente sont largement diffusés par divers sites internet de référence alors que mon contravis n'est cité qu'occasionnellement par des renvois à mes publications récentes. Je peux donc considérer que la circulation de mon point de vue se trouve fortement désavantagée.

Il y a un obstacle encore plus important à la poursuite d'un débat axé sur ce thème. En effet, comme je l'ai dit ci-dessus dans la présentation de cet exposé, je précise des principes que je veux généraux afin de montrer qu'ils ne permettent pas d'en tirer la conclusion que je conteste, autrement dit, j'affirme qu'il y a eu une erreur. Or, selon les critères actuels du discours scientifique, en faisant cette affirmation, je quitte le domaine du simple débat pour rentrer dans celui, souvent jugé stérile, de la polémique que les habitudes discursives de la communauté scientifique tendent à éviter.

Les normes imposées en ce sens sont clairement exposées par les règles de participation à Wikipédia, vd. https://en.wikipedia.org/wiki/Wikipedia:Ten_simple_rules_for_editing_Wikipedia¹⁰⁵ :
Treat other editors as collaborators and maintain a respectful and civil manner, even in disagreement.¹⁰⁶

Or, si je m'exprime comme je le fais ici, les auteurs des exposés mis en cause peuvent considérer que je leur manque de respect.

All articles in Wikipedia should be impartial in tone and content. When writing, do state facts and facts about notable opinions, but do not offer your opinion as fact.¹⁰⁷

Par conséquent, si je veux exprimer ma façon de comprendre l'essence de la linguistique historique, je devrais la présenter comme une possibilité parmi d'autres et la soupeser et cela n'a pas été fait ici.

Cette norme n'est pourtant pas toujours facile à respecter :

Occasionally you may interact with another editor who clearly does not share your expertise on the subject of an article. This can often prove frustrating...¹⁰⁸

Il suffirait dans ce cas d'apporter les références déterminantes :

Your specialist knowledge should enable you to write in a neutral manner and produce reliable, independent sources to support each assertion you make.¹⁰⁹

¹⁰⁵ L'équivalent français semble être https://fr.wikipedia.org/wiki/Wikipédia:Règles_de_savoir-vivre.

¹⁰⁶ Cf. la version française : [I] il est fortement recommandé de garder avec ses interlocuteurs un esprit de non-violence, et de communiquer avec eux avec respect et courtoisie.

¹⁰⁷ « Wikipédia n'est pas un forum de discussion. En conséquence, tous les participants doivent avant tout arriver à un consensus, afin de collaborer avec plus d'efficacité dans la construction de l'encyclopédie. Ainsi, pour rendre les articles aussi objectifs que possible, même les articles aux sujets controversés, ceux-ci ne doivent pas être le prétexte à un quelconque prêche. »

¹⁰⁸ Je n'ai pas trouvé l'équivalent en français de cette reconnaissance sur les pages consultées.

¹⁰⁹ En français, on est orienté vers la page de discussion pour créer le consensus : « Parfois, des désaccords naissent entre contributeurs de Wikipédia. La règle indispensable est donc de s'expliquer et de donner aux autres l'opportunité de faire de même sur les pages de discussion. » <<https://fr.wikipedia.org/wiki/Aide:Discussion>>

Ce texte a peut-être plusieurs interprétations. La mienne en est la suivante : la science n'est pas un avis, c'est l'expression de ce qui est. On reconnaît l'expression de ce qui est par le fait que les scientifiques sont d'accord pour le dire. La mise en minorité invalide le discours des tenants d'un avis quelconque. C'est toujours gênant lorsque quelqu'un refuse de se ranger à un avis majoritaire.

Cette façon de penser convient particulièrement au cadre d'une encyclopédie conçue comme le lieu où les savants disent aux non-savants ce qu'il faut penser. Si un auteur se range d'un côté d'une dispute présentée comme non encore résolue, le lecteur profane n'aura pas les moyens de savoir s'il doit lui apporter son soutien ou pas. Par conséquent, on déclare la résolution du conflit en suspens sans prendre parti. Toutefois, lorsque la majorité penche fortement d'un côté, on peut lui accorder le droit d'imprimer son point de vue : on trouvera dans la littérature un volume de références convainquant.

Dans cette perspective, considérons l'article « mass comparison » (comparaison multilatérale) chez Wikipédia en anglais¹¹⁰, particulièrement pertinent pour notre propos. Nous y trouvons dans les premières lignes les affirmations suivantes :

Mass comparison is a method developed by Joseph Greenberg to determine the level of genetic relatedness between languages. It is now usually called multilateral comparison. The method is rejected by most linguists (Campbell 2001¹¹¹, p. 45), though not all.

Some of the top-level relationships Greenberg named are now generally accepted, though they had already been posited by others (e.g. Afro-Asiatic and Niger–Congo). Others are accepted by many though disputed by some prominent specialists (e.g. Nilo-Saharan), others are predominantly rejected but have some defenders (e.g. Eurasiatic), while others are almost universally rejected (e.g. Khoisan and Amerind).

En totale conformité avec la règle, la présentation est neutre (« though not all ») tout en revendiquant la majorité à travers la citation d'un linguiste distingué et par conséquent le droit de prendre en charge la rédaction de l'article et de l'orienter. Fort de ce pouvoir, on passe immédiatement à dévaloriser le travail de Greenberg (« they had already been posited by others »...) en infirmant en même temps la position des « not all » qui lui sont toujours favorables.

Cette présentation a ses pierres d'achoppement : entre autres le fait de choisir comme « source indépendante » un américaniste qui a rendu compte de Greenberg (1987) en affirmant « the classification [in *Language in the Americas*] and its attendant methods must be rejected » (Campbell 1988 : 610) ainsi que l'absence de toute enquête permettant de déterminer avec certitude quel camp emporte réellement la majorité des suffrages¹¹². Toutefois, il n'est pas envisageable qu'un opposant intervienne en changeant des points particuliers de cette introduction ni des paragraphes « Relation to the comparative method » and « Criticism » qui la suivent. De cette façon, par un nuancement l'un des adversaires dans ce qui est en fait une polémique prend l'avantage.

¹¹⁰ https://en.wikipedia.org/wiki/Mass_comparison

¹¹¹ Le texte publié est Campbell (2003).

¹¹² Il serait d'ailleurs assez compliqué de savoir qui sont ceux qui ont suffisamment réfléchi sur la question pour pouvoir formuler un avis fondé. De plus, il est déconcertant de vouloir accepter ou rejeter en bloc une méthodologie.

Toutefois, au mérite de Wikipédia, l'équilibre est restauré par une solution *ad hoc* : on a simplement permis le rajout de deux paragraphes qui présentent un point de vue favorable à la comparaison multilatérale tout aussi nuancé. En effet, la discussion est élevée au plan d'un bilan contrasté des pour et des contre : d'un côté, on a la notion que c'est la régularité des correspondances phonétiques (avec explications des irrégularités en cas de besoin) qui prouve l'existence d'un rapport génétique entre langues, tout en imposant une limitation à la profondeur temporelle des reconstructions ; de l'autre, on constate l'existence de ressemblances forme/sens qui permettent de conclure à des rapports génétiques flous mais intéressants d'une ancienneté très supérieure.

Prima facie, on a ainsi obtenu la coexistence pacifique. On propose à ceux qui reconstruisent de reconstruire autant qu'ils peuvent et à ceux qui veulent voir plus loin de relever tous les rapprochements repérables, sans se soucier les uns des autres. Or, le fondement de la polémique subsiste : un désaccord sur la nature de la linguistique historique. En effet, en affirmant comme je le fais en §5 qu'en plus de restituer les composants phonologiques d'éléments lexicaux, cette discipline cherche à comprendre la structure de la protolangue et le fonctionnement de sa diachronie, je fais revivre la polémique en prenant position : j'affirme que le but de la linguistique historique ne se résume pas au fait de préciser le rapport génétique entre certaines langues mais vise en même temps l'amélioration de notre compréhension du langage humain, notamment du phénomène de changement permanent qui lui est inhérent¹¹³ : nous décrivons le réel afin de déterminer le domaine du possible.

Dans cette perspective je déplace ouvertement la dualité suggérée par les pro-Greenberguiens de l'article Wikipédia vers un autre plan en disant qu'on fait de la linguistique comparative afin de connaître les modalités du changement dans les langues et qu'on fait de la comparaison multilatérale afin d'identifier les langues qui partagent une histoire¹¹⁴. Si on accepte ce partage de rôles, on devrait pouvoir analyser sereinement les postulats de la comparaison multilatérale, par exemple, en évaluant l'utilité de la notion réductrice de « parenté génétique » qu'elle invoque et en catégorisant les configurations de ressemblances, tant phonétiques que sémantiques¹¹⁵, qu'elle met en jeu.

A mon avis, une évaluation critique de ce genre amènera à la conclusion que l'un des problèmes principaux posés par la comparaison multilatérale est son ambiguïté et l'incompréhension conséquente qu'elle génère. Considérons la notion de « reconstruction » ou de « protoforme ». Ce sont des termes qui peuvent désigner

► un élément antérieur à des changements phonétiques en cours (vd. l'anglais américain selon Labov *et al.* 2006) ;

¹¹³ Pour cela une compréhension des conséquences des contacts de langues sera aussi essentielle que la découverte de changements internes : nous visons un phénomène global.

¹¹⁴ Précisons toutefois qu'on ne fait ni l'une ni l'autre afin de connaître l'Histoire. C'est l'historien qui pourra se servir de l'une et/ou de l'autre afin d'éclairer les faits qu'il étudie et les résultats de ses recherches pourront aider le linguiste à mettre en perspective les circonstances dans lesquelles le changement linguistique et la diversification des langues se produisent.

¹¹⁵ La comparaison multilatérale prend-elle en compte des traits typologiques ? Si oui, de quel ordre ? Et est-ce une qualité ou un défaut ?

- ▶ un étymon du *Dictionnaire étymologique roman* (Buchi & Schweickard 2015) ;
- ▶ une forme proto-indo-européenne de chez Lehmann (1952) ;
- ▶ une unité restituée par Guthrie (1971) pour le bantou commun ;
- ▶ un élément proposé par Ehret (2001) pour le proto-nilosaharien ; et
- ▶ une forme composite comme celles auxquelles je fais référence ci-dessus (§4).

Il devrait être évident que des entités établies

- ▶ moyennant l'écoute d'enregistrements contemporains ;
- ▶ par l'étude de langues bénéficiant de témoignages écrits sur la quasi-totalité de leur histoire ;
- ▶ à partir de conjectures formulées sur la base d'écrits dont certains ont une ancienneté de plusieurs milliers d'années ;
- ▶ pour des langues proches mais dont l'histoire écrite prédate à peine le 20^e siècle ;
- ▶ pour des langues très différenciées et sans témoignage écrit dont la protolangue serait d'une grande ancienneté ;

▶ et par la simple agglomération de traits phonétiques trouvés dans des morphèmes jugés ressemblants,

ne peuvent pas être de même nature, même si elles sont toujours censées représenter par l'écrit une production sonore qui aurait pu être émise par un être humain à un moment donné du passé. Chaque linguiste peut choisir le domaine de la linguistique historique qui l'intéresse mais, s'il reconstruit, il ne doit pas supposer des équivalences entre ses reconstructions et celles qui sont proposées dans d'autres systèmes. Si on part de ce principe et qu'on laisse faire chacun, il n'y aura pas d'empiètement et on pourra évaluer sans énervement les méthodes et les résultats obtenus dans chaque contexte. Si on avait procédé ainsi après la publication de Greenberg (1987), Campbell (1988 : 610) n'aurait pas eu besoin de regretter que cet ouvrage « complicates the task of getting on with the business of working out American Indian linguistic relationships » car on n'aurait jamais eu l'idée de mettre sur le même plan les étymons supposés par Greenberg et, par exemple, les reconstructions de Stubbs (2011) pour le proto-uto-aztèque. Chaque entité est à juger d'abord pour ce qu'elle est en elle-même et non pas pour ce qu'elle représente à l'égard d'une autre.

Ce « mélange de genres » existe parce que, au lieu d'accorder aux résultats de la comparaison multilatérale le statut d'aboutissement, on voudrait en faire un point de départ, une esquisse ouvrant la voie à un raffinement sans fin de la définition des rapports de parenté entre langues en supposant que chaque nouvelle donnée peut valoir une restructuration. Or, on peut à la rigueur étoffer les classements de Greenberg : rajouter dans un groupe donné une langue qu'il n'avait pas prise en considération, trouver dans une langue donnée un représentant d'un étymon qu'il n'avait pas relevé parce que ce terme ne se trouvait pas dans sa documentation ou parce qu'il avait pris un sens différent, évaluer les groupes en termes de comparabilité intuitive en mettant de côté les langues qui ne seront pas intégrées aux premiers niveaux de reconstruction et, bien entendu, signaler des cas problématiques de langues qu'on a du mal à situer (vd. en Afrique le laal de Boyeldieu 1982) et envisager leur traitement. Mais notre connaissance de la diachronie des langues ne sera pas améliorée par le survol permanent afin d'extraire des descriptions un élément par ci, un autre par là. Le classement qu'on possède suffit largement afin d'orienter des études comparatives basées sur la

mise en commun des connaissances descriptives qui s'accroissent progressivement et c'est par de telles études que notre compréhension du fonctionnement des langues africaines en diachronie sera affinée.

13.2. La distillation

Mes objections au reclassement du sama se situent donc parfaitement dans ce cadre polémique. Le commentaire préliminaire de B-A :

Especialy in an area linguistically so fragmented as the Nigeria-Cameroun-Chad region (where the Adamawa-Eastern languages are concentrated), it is almost inevitable that individual languages should be mis-assigned. (B-A 24)

est un rejet de fait de la validité de la comparaison multilatérale et préfigure les critiques de Campbell (1987 : 610) qui intègre la réponse de Greenberg tout en voulant l'infirmier :

But G[reenberg] would claim insulation from problems in the data: 'The fact is, the method of multilateral comparison is so powerful that it will give reliable results even with the poorest of materials' ([1986 :]29. G's method is 'powerful', but it is not reliable, as we have seen here.

Ma contestation doit donc avoir deux composantes, d'un côté la réfutation du raisonnement utilisé par B-A pour justifier le reclassement du sama mais d'un autre côté une défense de la comparaison multilatérale qui valide l'utilité d'une unité classificatoire Ad-Ou où le sama sera maintenu. Le reclassement génétique du sama, étant un cas unique¹¹⁶, peut sembler être un détail de portée minimale mais il est l'affleurement d'un iceberg conflictuel bien plus vaste et doit être interprété à la lumière de la confrontation d'écoles des pour- et des anti-comparaison multilatérale.

Je suggère, comme Newman (1995) l'a fait bien avant moi, que ce conflit est fondé sur une mésinterprétation de la démarche de Greenberg par ses contradicteurs. Or, l'extension de ce phénomène est telle qu'elle exige en soi une explication : pourquoi autant de linguistes de compétence reconnue réagissent-ils négativement aux thèses de Greenberg ?

Pour ma part, je vois cette situation comme une conséquence d'un clivage encore plus profond qui détermine les orientations des linguistes, quels que soient leurs domaines de choix. On trouve d'un côté ceux pour qui le langage est une faculté dont ils participent en tant qu'êtres humains. L'étude d'une langue, en particulier d'une langue étrangère, consiste à exercer cette faculté : on cherche à rejoindre l'expérience du locuteur natif et on pense comprendre la langue lorsqu'on a un sentiment de l'avoir intégrée¹¹⁷. En face on a ceux qui pensent qu'une approche scientifique exige qu'on se détache de son objet d'étude et qu'on le réifie. Ils pensent avoir compris lorsqu'ils modélisent le comportement qu'ils observent en en donnant une description cohérente. Or, la question se pose : ces deux camps s'affrontent-ils toujours ? Ou peut-on se retrouver occasionnellement en compagnie des adversaires de son camp de prédilection ? Se voit-on parfois

¹¹⁶ Malgré sa déclaration de méfiance, B-A ne relève qu'une seule « erreur » : le sama. Un cas semblable concernant le tiv (glottocode tivv1240, ISO 639-3 tiv) s'est présenté chez B&S (1977). Cette langue partage avec les langues bantoues de l'est et du sud un terme symbolisé **ungwa* signifiant 'entendre'. Par conséquent, la priorité étant donnée aux « innovations » face au simple comptage, ces langues doivent être classées ensemble. Mais ce choix étant universellement rejeté par les spécialistes comme invraisemblable, le reclassement est abandonné contrairement au cas du sama.

¹¹⁷ Cf. mes remarques concernant l'intuition, §12.

dans des circonstances qui exigent une conversion et qu'il faut reconnaître comme telles lorsqu'elles apparaissent ? Je n'ai aucune référence documentaire où ces questions auraient été soulevées et explorées. Il se peut qu'elles soient trop immédiates pour être perçues et que cela entraîne un certain vide de réflexion. Quel que soit le cas, les linguistes ont certainement intérêt à prendre davantage conscience des manières dont leurs prédispositions déterminent leur comportement professionnel.

Références bibliographiques

- Bearth, Thomas & Christa Link. 1980. The tone puzzle of Wobe. *Studies in African Linguistics* 11(2) : 147-207.
- Bender, M. Lionel. 1997. *The Nilo-Saharan Languages*. LINCOM Handbooks in Linguistics 06. München : LINCOM.
- Bendor-Samuel, John. *The Niger-Congo languages*. Lanham : University Press of America.
- Bennett, Patrick R. 1983. Adamawa-Eastern: Problems and Prospects. In Dihoff, Ivan R. (éd.), *Current Approaches to African Linguistics*, 23-48. Dordrecht : Foris Publications.
- Bennett, Patrick R. & Jan P. Sterk. 1977. South Central Niger-Congo: a reclassification. *Studies in African Linguistics* 8 (3) : 241-273.
- Blench, Roger. 2006. Dong wordlist. <<https://www.rogerblench.info/Language/Niger-Congo/BC/EBCOP.htm>, « Dakaoid » >
- 2008, Prospecting proto-Plateau. <<https://www.rogerblench.info/Language/Niger-Congo/BC/Plateau/General/PIGOP.htm>, « Prospecting proto-Plateau » >
- 2014. The Bantoid languages <https://www.academia.edu/6026186/The_Bantoid_languages
- sans date1. The North Bantoid hypothesis <<https://www.rogerblench.info/Language/LingOP.htm>> Niger-Congo > Bantoid, « North Bantu Hypothesis » >
- sans date2. Niger-Congo reconstruction. <<https://www.rogerblench.info/Language/Niger-Congo/General/NCgenOP.htm>>
- Blench, Roger & Kay Williamson, 1987, 'New' Bantoid. MS.
- Bohnhoff, Lee. 2019. Dictionnaire dii (dourou) - français avec indexes français et anglais. Robert Hedinger & Marlene Ferguson eds. Kay Williamson Educational Foundation – African Languages Monographs 11. Köln : Rüdiger Köppe.
- Boyd, Raymond. 1989. Adamawa-Ubangi. In Bendor-Samuel, John, *The Niger-Congo languages*, Lanham : University Press of America, 178-215.
- 1994. *Historical Perspectives on Chamba Daka*. Köln : Rüdiger Köppe.
- 1995. De l'expression à l'expressivité en morphologie : analyse comparée de la dérivation verbale en zande et en nzakala, *Afrikanistische Arbeitspapiere* 43: 5-36.
- 1996/1997. Chamba-Daka and Bantoid: a further look at Chamba-Daka classification. *Journal of West African Languages* 26(2):29-43.
- 1999. A linguistic sketch of Tiba (Ga), Part I, *Afrika und Übersee* 82:1-17 ; Part II, *Afrika und Übersee* 82:213-49.
- 2004. The syntax and semantics of the Chamba-Daka verbal noun. *Afrika & Übersee* 87: 193-288.
- Boyd, Raymond & Helma Pasch. 1988. Le groupe sere-ngbaka-mba. In Yves Moñino, (ed.), *Lexique comparatif des langues oubanguiennes*. Paris: Geuthner, 51-58

- Boyeldieu, Pascal. 1982. *Deux Etudes laal (Moyen-Chari, Tchad)*. Berlin: Reimer.
- Boyeldieu, Pascal & France Cloarec-Heiss. 1986. Dialectométrie lexicale dans le domaine oubanguien. In Gladys Guarisma & Wilhelm J. G. Möhlig (éds). *La méthode dialectométrique appliquée aux langues africaines*. Berlin : Dietrich Reimer, pp. 331-393.
- Boyeldieu, Pascal & Raimund Kastenholz & Ulrich Kleinewillinghöfer & Florian Lionnet. 2020. The Bua group noun class system: looking for a historical interpretation. In Sabine Littig & Friederike Vigeland & Alexander Zheltov (eds.), *Proceedings of the first Adamawa Conference (Mainz, September 9–11 2019)*. Moscow: Inst. Jazykoznanija, RAN, 181–215.
- Buchi, Eva & Wolfgang Schweickard. 2015. Dictionnaire Étymologique Roman (DÉRom) : Genèse, méthodes et résultats. Berlin / Boston : De Gruyter.
- Campbell, Lyle. 1988. Review of Greenberg, *Language in the Americas*. *Language* 64(3) : 591-615.
- . 2003. Beyond the Comparative Method?. In Blake, Barry J.; Burridge, Kate; Taylor, Jo. (eds.). *Historical Linguistics 2001. Selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001*. Amsterdam/Philadelphia : Benjamins, 33-58.
- Cloarec-Heiss, France. 1998. Entre Oubanguien et Soudan central : les langues banda. In I. Maddieson & Th. J. Hinnebusch (éds), *Language History and Linguistic Description in Africa*. Trenton-Asmara : Africa World Press (Trends in African Linguistics 2), 1-16.
- . 2000. Mesures dialectales en trois dimensions : application à une aire dialectale hétérogène, l'aire banda. In Wolff, H. E. & G. Gensler (éds), *Proceedings of the second world congress of African linguistics, Leipzig 1997*. Cologne : Rüdiger Köppe, 175-195.
- Connell, Bruce. 2000. The integrity of Mambiloid. In H. E. Wolff & O. D. Gensler (eds). *Proceedings of the 2nd World Congress of African Linguistics*. Köln : Rüdiger Köppe, 197-213.
- Dieu, Michel & Louis Perrois avec la collaboration d'Henry Tourneux. 2016. *Dictionnaire encyclopédique koma g'imbē / français (Monts Alantika, Nord-Cameroun)*. Studi africanistici : Serie Ciado-Sudanese 8. Napoli : Università degli studi di Napoli "L'Orientale".
- Ehret, Christopher. 2001. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln : Rüdiger Köppe.
- Fabre, Gwenaëlle. 2003. *Étude du samba leko, langue Adamawa du Cameroun*. Lincom Studies in African Linguistics 56. Munich : Lincom.
- Fardon, Richard. 1988. *Raiders and refugees : Trends in Chamba political development, 1750-1950*. Washington : Smithsonian.
- Greenberg, Joseph. 1957. *Essays in Linguistics*. Chicago : University of Chicago Press.
- . [1963] 1966. *The languages of Africa*. Bloomington: Indiana University.
- . 1974. Reply to Prof. Meeussen. *Studies in African Linguistics*, Supp. 5 : 122-4.
- . 1987. *Language in the Americas*. Stanford : Stanford University Press.
- Guarisma, Gladys. 1986. In Gladys Guarisma & Wilhelm J. G. Möhlig (éds). *La méthode dialectométrique appliquée aux langues africaines*. Berlin : Dietrich Reimer, pp.
- . 2000. *Complexité morphologique, simplicité syntaxique : le cas de bafia, langue bantoue périphérique (A50) de l'Ouest du Cameroun*. Paris : Peeters, Langues et cultures africaines 24, SELAF 379.

- Güldemann, Tom. 2019. Areal linguistics in Africa before a new approach to its genealogical language classification: 1. African language classification beyond Greenberg. Conférence délivrée le 9/3/2019 au LLACAN, Paris.
- . 2021. From asymmetric grammatical treatment of nouns to gender: animacy-based noun classification in Central Africa and its typological significance, ms.
- Guthrie, Malcolm. 1971. *Comparative Bantu*. Farnborough : Gregg Press.
- Hedinger, Robert. 1989. Northern Bantoid. In John Bendor-Samuel, éd, *The Niger-Congo languages*, Lanham : University Press of America, 421-429.
- Husserl, Edmund. 1986. *Méditations cartésiennes*. Paris : Vrin.
- Kleinwillinghöfer, Ulrich. 1996. Relationship between Adamawa and Gur languages: The case of Waja and Tula. *Cahiers Voltaïques - Gur Papers*, 1: 25-45.
- . 2011. Bəna-Mboi or Yungur Group. < <https://www.blogs.uni-mainz.de/fb07-adamawa/adamawa-languages/bəna-mboi-yungur-group/> >
- . 2014. The languages of the Tula-Waja Group. < <https://www.blogs.uni-mainz.de/fb07-adamawa/adamawa-languages/tula-waja-group/> >
- Labov, William; Ash, Sharon; Boberg, Charles. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin / New York: Mouton de Gruyter.
- Lehmann, Winfred P. 1952. *Proto-Indo-European Phonology*. Austin: University of Texas Press and Linguistic Society of America.
- Manessy, Gabriel. 1975. *Les langues oti-volta : Classification généalogique d'un groupe de langues voltaïques*. Paris : SELAF.
- . 1979. *Contribution à la classification généalogique des langues voltaïques : le groupe proto-central*. Paris : SELAF.
- . 1992. Généalogie et génétique. *Linguistique africaine* 9 : 67-75.
- Martin, Marieke. sans date. https://www.academia.edu/36939635/Noun_class_remnants_and_number_systems_in_Mambiloid.
- Newman, Paul. 1995. *On being right*. Bloomington : West African Languages Institute.
- Person, Yves. 1968. Des Kru en Haute-Volta (Sème ou Siamou de Orodara). *Bulletin de l'IFAN* 28B (1-2) : 485-492).
- Schuchardt, Hugo. 1928 (2e éd.). *Hugo Schuchardt-Brevier: Ein Vademecum der allgemeinen Sprachwissenschaft*. Leo Spitzer, éd. Halle/Saale: Niemeyer.
- Shimizu, Kiyoshi. 1979. *A comparative study of the Mumuye dialects (Nigeria)*. Berlin : Dietrich Reimer.
- Stewart, John. 2002. The potential of Proto-Potou-Akanic-Bantu as a pilot Proto-Niger-Congo, and the reconstructions updated, *Journal of African Languages and Linguistics* 23(2) : 197-224.
- Strümpell, Kurt. 1910. Vergleichendes Wörterverzeichnis der Heidensprachen Adamawas. *Zeitschrift für Ethnologie* 42 : 444-488.
- Stubbs, Brian D. 2011. *Uto-Aztecan: A comparative vocabulary*. Flower Mound, Texas: Shumway Family History Services.
- Vogler, Pierre. 2019. Le sèmè/siamou n'est pas une langue kru. < <https://hal.archives-ouvertes.fr/hal-01182225v2> >