

Les lexiques proto-bantous

De quel matériau dispose-t-on quand on aborde les langues bantoues sous l'angle de la linguistique historique ? Jusqu'à ces dernières années, les études comparatives s'appuyaient sur deux ouvrages majeurs : *Comparative Bantu* (C.B.) de M. Guthrie (1967-71) et *Bantu Lexical Reconstructions* (B.L.R.) de Meeussen (1969). Plus récemment, l'équipe de Tervuren a élaboré une première version informatisée des reconstructions sous l'appellation : B.L.R.2 (1998), puis B.L.R.3 (2003), qui s'entendent comme une suite du travail entrepris avec Meeussen.

Les données contenues dans C.B. et dans B.L.R.1 seront systématiquement analysées dans cette étude, d'où la nécessité de préciser comment chaque ouvrage a été élaboré. Étant donné, d'une part la complexité méthodologique de C.B. qui a engendré trois lexiques de niveau de profondeur différent, et d'autre part le fait que les études comparatives aient privilégié les Cs au détriment du PB X et de B.L.R.1, je m'attarderai un peu plus longuement sur la présentation de C.B.

1 – Comparative Bantu

L'ouvrage de Guthrie, paru sous ce titre, comprend quatre volumes dont deux sont consacrés à ce que cet auteur nomme le bantou commun, c'est-à-dire un ensemble de cognats à partir desquels il a extrait la langue-mère : le proto-bantou. Il applique les principes de la démarche hypothético-déductive où il ne peut y avoir inférence qu'à partir de données vérifiables. Dans ce cadre, la première étape a été la constitution du bantou commun, la deuxième étape étant une étude de la préhistoire bantoue à partir des données du bantou commun.

1.1 - Le bantou commun

Le bantou commun rassemble des listes de cognats appelées séries comparatives (Cs), et représentées par une forme à astérisque ("starred form"). Ces séries comparatives sont fréquemment utilisées dans les études comparatives au même

titre que des reconstructions alors qu'elles n'en sont pas. On verra dans la suite de cette présentation des travaux de Guthrie que ses reconstructions, au sens strict du terme, appartiennent à ce qu'il nomme le PB X.

1.1.1 - Les séries comparatives

Les 2300 séries comparatives (Cs) rassemblent autour d'un sens commun les cognats recueillis à travers 300 langues bantoues, en sachant que pour une partie des langues, dont il ne précise ni le nombre ni le nom, les données dont dispose Guthrie, sont très partielles. Dans cette première étape, Guthrie s'appuyant sur sa "Classification of the Bantu Languages"¹ de 1948, suit les principes de la méthode comparative en établissant des listes d'items qui doivent répondre au critère de correspondance régulière. Pour Guthrie, les changements de sons produisent invariablement des correspondances régulières, pour Greenberg ces deux niveaux ne sont fortement corrélés que lorsque les langues sont très proches.

1.1.1.1 - Conditions de validation d'une Cs

Plusieurs conditions sont requises pour considérer comme valide une Cs : chaque Cs doit contenir au moins 3 items provenant de 3 zones différentes, et chaque correspondance doit être attestée au moins 3 fois dans chaque langue. Guthrie insiste sur la nécessité de ne retenir que des correspondances complètement régulières. Les correspondances sont testées par syllabe : C_1V_1 et C_2V_2 pour les noms -CVCV ; C_1V_1 et V_1C_2 pour les verbes -CVC-.

1.1.1.2 - Séries partielles

Lorsqu'une liste de correspondances est insuffisamment représentée, c'est-à-dire quand elle ne répond pas aux conditions décrites ci-dessus, elle figure dans le bantou commun sous l'appellation de série partielle (ps). Sous le terme de ps, on

¹ **Classification des langues bantoues** : une des premières publications de Guthrie a porté sur la classification des langues bantoues (1948). Il a divisé l'ensemble du domaine bantou en 15 zones, selon des critères géographiques, et a attribué à chacune de ces zones une lettre de A à S. Chaque zone comprend plusieurs groupes de langues, dont la proximité linguistique est variable. Au cours des années 80, Meeussen a isolé une nouvelle zone ; la zone J, ou zone interlacustre, qui regroupe des langues des zones D et E. (cf.colloque de Viviers) Une fois ces zones définies, Guthrie les a regroupées en 6 aires qui lui permettront de préciser le degré de dispersion géographique de ses séries comparatives : • Nord-Ouest (NW) = zones A, B, C, • Centre- Ouest (CW) = zones H et L, • Sud-Ouest (SW) = zones K et R, • Nord-Est (NE) = zones D, E et F, • Centre-Est (CE) = zones G, M et N, • Sud-Est (SE) = zones P et S.

trouvera aussi bien des listes de cognats ayant une répartition géographique insuffisante, car provenant de langues n'appartenant pas à trois zones distinctes, que des listes provenant de langues nettement éloignées mais limitées à deux cognats. Ainsi des listes de cognats, à distribution régionale, reçoivent le même traitement que des listes de cognats provenant de deux langues géographiquement éloignées. On verra dans l'analyse du corpus que certaines séries partielles se retrouveront au niveau de la proto-langue.

1.1.1.3 - Séries "osculant"

Un certain nombre de Cs, semblables du point de vue du sens, diffèrent plus ou moins du point de vue de la forme, alors que d'autres, semblables du point de vue de la forme, diffèrent plus ou moins du point de vue du sens. Ainsi, la moindre variation a entraîné la constitution d'une nouvelle Cs. Guthrie qualifie dans un premier temps ces Cs d'osculant et prendra en compte ce qualificatif lorsqu'il recherchera les Cs susceptibles d'appartenir au proto-bantou. La définition de ce concept, qu'il a emprunté aux mathématiques, évoquerait une situation de variation mais c'est une notion qu'il n'utilise pas.

1.1.1.4 - Multivalence

Un dernier point me semble important à préciser, il concerne l'utilisation que fait Guthrie du concept de multivalence. Pour une langue donnée, un cognat est dit multivalent lorsqu'il peut appartenir à plusieurs Cs. Les cognats multivalents sont très présents dans les listes à initiale palatale et ont amené Guthrie à poser des formes à astérisque différentes, malgré des listes quasi identiques, qui le conduisent à opposer *y à *j ou *y à *g. Dans le cas présent, la notion de multivalence pointe indirectement les hésitations de Guthrie, quant à la qualité du segment à reconstruire.

1.1.1.5 - Les "starred forms" ou formes à astérisque

Une fois les listes de correspondances établies, chaque Cs est symbolisée par une forme à astérisque, qui n'est pas une reconstruction mais "*une représentation symbolique des rapports entre les contours phonématiques des divers membres d'une série*" (Guthrie 1959).

• Comment décide-t-il de la forme de cette représentation ?

Comme pour établir les correspondances, il examine par syllabe les cognats et pose pour chaque syllabe la forme qui permettra de rendre compte de ce qui a été observé dans les langues. Mais ce travail a été effectué par étapes successives puisqu'il précise (§41.22) que dans un premier temps : "*there is a careful selection of*

the data, in which the more complicated patterns are temporarily passed over". Ensuite, la première "starred form" obtenue est testée avec le reste des données et éventuellement modifiée. Chaque syllabe est construite avec des segments réalisés dans les langues actuelles. Son objectif est de rester au plus près de la synchronie. Les formes à astérisque portent un ton mais les cognats dont elles dépendent n'ont aucune indication tonale.

- **Attribution du sens**

Le sens attribué est celui qui, commun à la majorité des cognats, a permis de les réunir dans la même liste, la moindre nuance a déjà entraîné la construction d'une autre liste de cognats.

- **Localisation et dispersion**

Chaque "starred form" est accompagnée de deux précisions relatives à la situation géographique des cognats qui composent la Cs. D'une part, est indiqué le coefficient de dispersion de chaque Cs, calculé à la fois sur le nombre de zones représentées, par rapport à l'ensemble des 15 zones formant le domaine bantou, et sur leur degré de contiguïté vs éloignement. D'autre part, est précisée la localisation aréale des cognats, en fonction des six aires qu'il avait préalablement, définies (nord-ouest ; centre-ouest ; sud-ouest ; nord-est ; centre-est ; sud-est). Ces éléments, d'ordre géographique, ne sont indiqués que pour les données lexicales, non pour les données morphologiques. Nous verrons au paragraphe suivant, que ces caractéristiques seront déterminantes dans la démarche de Guthrie.

1.2 - Le proto-bantou

Le passage du statut de "starred form" à celui de **reconstruction** se fera exclusivement en fonction des critères géographiques et ce sont les 24% de Cs ayant une **distribution générale** qui formeront le proto-bantou X (PB X) et recevront le terme de reconstruction. Deux types d'attribution sont possibles pour qu'une forme à astérisque soit considérée comme générale.

- **Attribution directe**

On a vu précédemment que chaque Cs étant qualifiée du point de vue de sa répartition aréale. Lorsqu'elle est étiquetée GG ou G signifiant qu'elle a des cognats dans chacune des quinze zones, ou dans la plus grande partie du domaine bantou, alors elle sera considérée comme générale et reflétant directement la proto-langue.

- **Attribution indirecte**

Un groupe de plusieurs Cs "osculant" peut également être considéré comme général, lorsqu'en ajoutant leurs différentes aires de dispersion (joint spread) une distribution générale est obtenue. La notion de distribution générale est donc attribuée, soit directement par comptage du nombre de zones représentées dans la liste de cognats, soit après étude des séries "osculant".

• **Choix du proto-item**

Les séries "osculant" posent à Guthrie un certain nombre de problèmes, entre autres comment aboutir à un seul proto-item et comment le choisir ? En effet, il cherche à poser un seul proto-item et à lui subordonner les autres, en faisant appel à la notion de réflexe indirect. Il parle de mutation pour opposer ces changements fortuits aux changements réguliers. Il propose alors, de prendre en compte la direction de changement la plus probable et, à nouveau, la distribution géographique de chacune des Cs appartenant au groupe. Quand direction et distribution s'opposent, il semble privilégier la direction (ex Cs 444/507). Dans certains cas, il justifie l'existence des Cs synonymes en les qualifiant de réflexes indirects, dont la présence serait due à la mutation d'un item source secondaire et il en vient à postuler une étape PB X² intermédiaire entre PB X et PB A et B, qui focaliserait les innovations (§84.22).

D'autre part, Guthrie suit un présupposé de simplification. Lorsqu'une Cs à distribution générale présente des alternances de type -CVC-/-CVCVC-, il retient pour la proto-langue la forme -CVCVC-. Il n'envisage jamais la possibilité de lexicalisation d'un morphème grammatical (cf. CS 760, 995), processus pourtant répandu dans l'évolution des langues. Si la tendance universelle va plutôt dans le sens d'une érosion syllabique, et conduit les linguistes à reconstruire des formes lexicales plus longues, ce chemin n'est pas le seul emprunté par les langues et les processus de lexicalisation sont également à prendre à compte. Guthrie a appliqué de manière rigide ce qui n'est qu'une tendance.

2 – Travaux de Meeussen et de Tervuren

En 1969, Meeussen publie « Bantu Lexical Reconstructions » (B.L.R.) qui, contrairement au bantou commun, rassemble les reconstructions proposées antérieurement, par Meinhof, Homburger, Bourquin et Greenberg. Je rappellerai que l'on doit les premières reconstructions tonales du proto-bantou à Greenberg (1948). Meeussen a corrigé certaines reconstructions lexicales, leur a attribué un

²Le PB X² est représenté par une vingtaine de reconstructions dans l'appendice 8/2.

ton, et en a proposé de nouvelles. Le travail de Meeussen s'appuie sur les principes de la méthode comparative avec comme apport majeur le recours à la morphophonologie, à la notion de correspondance indirecte, donc à un niveau d'abstraction qui lui a permis des avancées significatives en matière de reconstruction. Le paragraphe est court, disproportionné par rapport à la section précédente consacrée à C.B., mais n'enlève rien à la qualité du travail contenu dans B.L.R.1.

Depuis 2003, l'équipe de Tervuren a mis à disposition de la communauté linguistique une version informatisée des reconstructions sous l'appellation de B.L.R.3. B.L.R.3 compile l'ensemble des reconstructions existantes qu'elles soient validées ou refusées. L'apport fondamental réside dans la mise en évidence des liens qui rattachent des items à un même étymon. Plusieurs niveaux de profondeur cohabitent dans B.L.R.3, et ces niveaux sont immédiatement identifiables.

Je vais résumer l'étiquetage utilisé dans B.L.R.3 pour qualifier un item car nous y aurons régulièrement recours dans cette étude. Les reconstructions au sens strict sont appelées MAIN. D'autres items leur sont rattachés par des étiquettes qui précisent leur lien de dépendance. Quand des reconstructions sont dérivées par un processus d'affixation -avec ou sans changement de catégorie grammaticale-, ou par un glissement sémantique elles sont rattachées à la reconstruction main par l'étiquette DER. Si l'étymon présente des variations formelles qui ne rentrent pas dans le cadre des correspondances régulières, il recevra l'étiquette VAR. D'autres sont COMPosés de plusieurs racines et rattachés aux reconstructions concernées. Dernier cas de figure, un étymon peut être REFusé au profit d'une autre reconstruction. Tout étiquetage établit un lien entre les reconstructions.

Une dernière catégorie est reliée à aucune reconstruction, il s'agit d'étymons isolés en attente de données complémentaires.