



**HAL**  
open science

## A propos d'un dictionnaire shingazidja-français

Michel Lafon

► **To cite this version:**

Michel Lafon. A propos d'un dictionnaire shingazidja-français. *Current Approaches to African Linguistics*, 1990, 7, pp.73-87. halshs-00267528

**HAL Id: halshs-00267528**

**<https://shs.hal.science/halshs-00267528>**

Submitted on 27 Mar 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A propos d'un dictionnaire shingazidja-français

Michel Lafon, CNRS - INALCO



La lexicographie bantou a toujours posé problème: les lexicographes hésitent entre une organisation par mots et un regroupement dans un seul *article* des mots de même famille ou apparentés, formés autour d'un même lexème.<sup>1</sup> L'organisation par mots est généralement jugée plus simple d'usage; ce fut celle retenue dans les premières descriptions de ces langues par des Européens; elle a la préférence en règle générale des locuteurs. Le regroupement par lexème, par contre, montre mieux le fonctionnement morphologique de la langue et permet d'en prendre en compte la productivité et la créativité: les processus de formation des mots étant décrits, le sens des termes non répertoriés pourra en principe être déduit dès lors que le lexème est présent, sous une forme ou sous une autre. Cette méthode a la faveur des linguistes qui, le plus souvent, ne sont pas des locuteurs natifs et est à l'honneur, avec plus ou moins de rigueur, dans la plupart des dictionnaires de langue bantou depuis la fin de la guerre<sup>2</sup>. C'est celle que nous avons adoptée, de façon systématique, dans la constitution du dictionnaire de shingazidja, en y ajoutant, pour faciliter la consultation, des renvois intégrés dans le corps du texte pour tous les termes n'apparaissant pas à leur place alphabétique. Insistons toutefois sur le fait que ces deux méthodes ne s'excluent nullement mais répondent chacune à des besoins et des motivations différentes.

## 1. Le shingazidja: situation linguistique et contexte culturel

Le shingazidja (grand-comorien) appartient aux parlers bantou de l'archipel des Comores, qui comprend quatre îles situées dans le canal de Mozambique, au sud de Zanzibar, à mi-distance de Madagascar et de la côte est-africaine. Le comorien se répartit en deux groupes entre lesquels l'intercompréhension n'est pas immédiate, shingazidja d'une part, anjouanais-mahorais de l'autre, le mohélien étant en quelque sorte intermédiaire. Le shingazidja quant à lui est parlé sur l'île de Grande-Comore (Ngazidja).

---

Je tiens à remercier pour leurs remarques sur une version antérieure de cet article Ch. Seydoux et les éditeurs. Cette présentation se réfère à Lafon 1987 et particulièrement au deuxième volume *Essai de dictionnaire étymologique*, dont nous préparons actuellement la publication d'une version augmentée.

Nous appelons article ce qui forme, dans cette optique, une unité lexicographique: un article comprend tous les termes issus d'un même lexème (noms, verbes, adjectifs, etc.); un article peut bien sûr n'être constitué que d'un terme. L'entrée (ou mot vedette) est le terme qui commande le classement alphabétique d'un article dans le dictionnaire, c'est-à-dire le premier terme de la liste.

Après une récapitulation "d'un siècle de lexicographie bantou", Benson (1964: 81-82), auteur d'un dictionnaire kikuyu où il met en pratique ce qu'il préconise, conclut "one cardinal principle which emerges from our study is that everything which needs to be said about a stem or root should be channelled into one single full article". Certains auteurs ont en outre introduit des distinctions graphiques entre ce qui est lexème et ce qui est affixe (au moyen de majuscules, d'italiques, de gras, etc.). Cela ne nous a pas semblé praticable dans le cas du shingazidja, du fait des amalgames fréquents en initiale de mot (voir plus loin); on ne peut, en effet, faire le départ, à partir des seules formes réalisées, entre ce qui est lexical et ce qui est grammatical; enfin, cela n'aurait pas été conforme à notre optique explicite d'instrumentalisation de la langue (voir note 5).

Les parlers comoriens appartiennent au groupe C40 de Guthrie, où figure aussi le swahili avec lequel ils partagent — du moins avec ses "dialectes primaires"<sup>3</sup> pour reprendre l'expression de Trevor Hill — la même situation socio-culturelle. La langue est issue de contacts, sans doute dès VII<sup>e</sup> siècle, entre des populations africaines bantu et des populations arabophones islamisées originaires de la côte sud-arabique et du Golfe<sup>3</sup>, elle est devenue la langue d'une société entièrement musulmane, où le prestige de l'Islam est très fort. Quoique indéniablement bantu, elle relève dans de nombreux domaines cette dualité qui se retrouve aussi sur le plan culturel. Ainsi le shingazidja comprend-il de nombreux emprunts à l'arabe qui ont enrichi le lexique mais aussi réagi sur les systèmes phonique et grammaticaux de sorte que toute étude ne peut éviter de prendre en compte cette hétérogénéité.

Les contacts entre les Comores et la côte sud-arabique se sont maintenus jusqu'à nos jours grâce aux relations commerciales, familiales, culturelles et, dans une moindre mesure, politiques (jusqu'au début du XX<sup>e</sup> siècle, les Comores ont plus ou moins été dans l'orbite du sultanat de Zanzibar — ce qui a d'ailleurs entraîné une influence du swahili lui-même aux Comores). Les relations avec les pays du Golfe et les pays arabes en général tendent actuellement à se renforcer: financement de projets de développement, appartenance des Comores à la Ligue Arabe, choix de l'arabe comme deuxième langue nationale après le français, etc. Sur le plan éducatif, l'école coranique traditionnelle, passage obligé de la très large majorité des enfants comoriens, dispense depuis longtemps une certaine connaissance de l'arabe et est maintenant relayée par l'enseignement de l'arabe comme matière dans les lycées ainsi que par la création, récente, d'une filière scolaire en arabe, parallèle au système utilisant le français.

Ainsi l'emprunt à l'arabe de nouveaux termes est-il généralement bien accepté, l'entrée de ces termes, préparée par les emprunts déjà présents, ne posant aucun problème d'intégration phonique ou morphologique. L'arabe sert de réserve lexicale dans une situation où il est prestigieux et où les emprunts à l'arabe, contrairement à ceux au français par exemple, sont moins ressentis comme extérieurs, culturellement sinon linguistiquement — de nombreux Comoriens, en effet, identifient la culture comorienne à la culture arabe ou musulmane et n'ont pas toujours conscience, en l'absence de toute éducation en langue vernaculaire, de l'autonomie du comorien par rapport à l'arabe (voir Laton à *paraître*).

Il existe ainsi un niveau de langue "arabisé", que l'on pourra identifier à celui des lettrés, qui comprend un grand nombre d'emprunts à l'arabe dans une prononciation volontairement étymologisante. Ce sociolecte est en principe doté d'un grand prestige et joue un rôle de référence parmi une partie de la population, notamment dans les campagnes.

A l'heure actuelle, le comorien ne bénéficie pas d'un statut reconnu, les langues officielles de la République Fédérale Islamique des Comores étant, comme il a été dit, le français et l'arabe; mais le comorien commence à être

<sup>3</sup>Les sites les plus anciens découverts aux Comores remontent, pour l'instant, au VI<sup>e</sup> ou VII<sup>e</sup> siècle (Mohéli); ils font apparaître des réseaux commerciaux reliant les rivages de l'océan indien (voir Chanutet 1988).

utilisé dans l'alphabétisation des adultes et des textes en comorien sont édités par des instances gouvernementales<sup>4</sup>.

## 2. Etat de la lexicographie shingazidja

Il n'existe pas de dictionnaire ni de lexique spécifique et récent du shingazidja. Les ouvrages existant montrent les deux types d'organisations, par racine et par mot: cions, de Sacloux le dictionnaire des dialectes swahili, recueilli au début du siècle, qui s'inspire d'un classement par lexème, où le shingazidja ne figure que marginalement, ainsi qu'un dictionnaire par mots de l'ensemble des parlers comoriens, qui est une compilation posthume à partir de ses notes et des indications contenues dans le premier dictionnaire (Ahmed-Chamanga et Gueunier 1979) et datant des années quarante, le lexique par mots figurant dans la grammaire-dictionnaire de Fisher 1949, de peu d'intérêt. Vient par contre de paraître un dictionnaire mahorais-français et français-mahorais de Blanchy 1987, organisé par lexèmes avec des renvois, donnant des indications étymologiques et l'on annonce un gros dictionnaire anjouanais-anglais de H. Ottenheimer.

Nous présentons ici la problématique qui a présidé à l'élaboration d'un dictionnaire shingazidja-français, organisé systématiquement par lexèmes, le regroupement des termes s'appuyant pour les emprunts à l'arabe sur l'étymologie. Nous exposons d'abord les modes de formation des mots, qui justifient la méthode lexicographique choisie, puis la constitution des articles de dictionnaire et les problèmes qui en découlent en nous cantonnant à ceux qui ont trait à la constitution formelle des articles — nous n'abordons pas ici le contenu ni l'organisation des définitions. Dans chaque cas, il convient de distinguer selon l'origine, bantu ou arabe, des mots.

## 3. Structure des mots ou constituants syntaxiques

### 3.1 Modèle bantu

Dire d'une langue qu'elle appartient à la famille bantu signifie notamment, sur le plan morphologique, que i) les constituants syntaxiques admettant l'accord de classes<sup>5</sup> s'analysent en une partie lexématique, porteuse de sens, qui peut être facultativement complétée par un ou plusieurs éléments appelés extension, et un ou plusieurs affixe(s) actualisateur(s), dont l'un au moins est obligatoire, et que ii) la même partie lexématique est susceptible de se

<sup>4</sup>Jusqu'à l'indépendance en 1975, le français était la seule langue utilisée officiellement dans l'archipel, à la notable exception de la loi de référendum d'auto-détermination, qui fut l'occasion d'une transcription, de la part des autorités françaises, en comorien en graphie arabe — discutée dans M. Ahmed-Chamanga et N-J Gueunier 1977. Après l'indépendance, le Gouvernement révolutionnaire d'Ali Soilih (1975-1978) tenta de promouvoir l'utilisation officielle du comorien en graphie latine en s'inspirant entre autres du travail fait à cette époque par Laron et Sibeitin-Blanc 1976. Cette option fut abandonnée après le changement politique de 1978 et toute utilisation officielle du comorien, proscrite. Ce n'est que depuis peu que l'on constate un changement d'attitude de la part des autorités.

<sup>5</sup>Le terme de constituant syntaxique est emprunté notamment à Houis 1983. Pour la terminologie et la problématique bantu, nous nous appuyons sur les travaux de Guthrie et la synthèse qu'en a donnée en français Alexandre 1981. Cet auteur oppose "les éléments participant à l'accord de classe et ceux qui n'y participent pas (la plupart sont de structure monomorphématique

retrouver dans plusieurs constituants syntaxiques présentant un certain rapport sémantique; on pourrait ajouter qu'une partie au moins de ces lexèmes doit se rattacher à un inventaire de racines dites communes, selon des correspondances régulières.

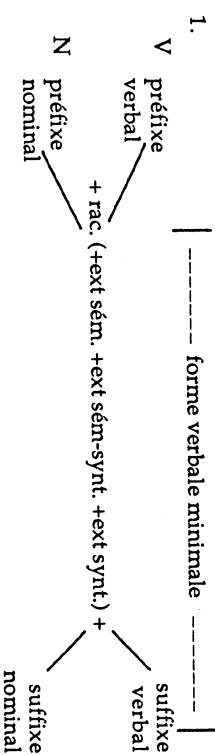
Les lexèmes en shingazidja se dissocient selon leur construction en deux grandes catégories:

- des racines bivalentes verbo-nominales qui admettent diverses dérivations (extensions)
- des thèmes monovalents nominaux qui n'admettent pas en principe d'extension<sup>6</sup>

Cela est illustré par les deux schémas suivants qui s'appliquent bien entendu aux mots du fonds bantou, ainsi que le montrent les exemples donnés ci-dessous, mais aussi aux termes d'emprunt "naturalisés", comme nous le verrons.

### 3.1.1 Constituants syntaxiques issus de racines

Les racines verbo-nominales donnent des verbaux et des nominaux:



Les barres verticales indiquent des formes structurales; pour rendre leur bivalence verbo-nominale, les racines sont glosées l' notion de ... 1. La racine primaire (ou nue) correspond à une racine dépourvue d'extension, alors qu'une racine étendue<sup>7</sup> est une racine munie d'une extension. La forme verbale minimale en (1) correspond à l'impératif 2<sup>e</sup> pers. sg. Nous distinguons des extensions sémantiques (2a), qui modifient le sens de la racine, des extensions syntaxiques (2b), qui impliquent en général un autre participant à l'énoncé et des extensions sémantico-syntaxiques (2c), qui combinent ces deux valeurs. (Voir Lafon 1987, vol. I: 157 ff.)

- 2a. extension sémantique      -t-
- réversible
- b. extensions syntaxiques:      -t- / -e-
- applicatif                    -(i)s- / -(i)z- / -(i)sh-
- causatif                        -ili- / -ele-
- intensif
- réciproque                    -an-
- c. extensions sémantico-syntaxiques
- passif                            -(i)w- / -(e)w-
- moyen-neutre                -(i)h- / -(e)h-

insécable" (Alexandre 1981: 362). Ces derniers termes, qui ne représentent qu'une faible minorité du lexique, ne nous concernent pas ici.

<sup>6</sup>Quelques thèmes nominaux dépendants (adjectivaux) admettent en fait des dérivations verbales.

Dans les exemples qui suivent, nous utilisons, à une exception près, le système orthographique préconisé par Ahmed-Chamanga et al 1988. Nous repreneons ci-dessous les symboles utilisés dans une valeur inhabituelle:

## 3a. graphes

b	= [b] (implosive)
d	= [d'] (implosive)
mC	nasale syllabique; deux syllabes: [ m <sup>u</sup> C ]
mC <sub>2</sub> nC	prénasalisées [ ~ C ] <sup>7</sup>
pV	[ β ] <sup>8</sup>
tr	[ t̥ ] (post-alvéolaire ou rétroflexe)

## b. symboles limités aux mots d'origine arabe

th	[ θ ]
dh	[ ð ]
gh	[ γ ]
kh	[ χ ]
ˀ	(apostrophe), attaque forte d'une voyelle
ˁ	(accent circonflexe) nasalisation

sur une voyelle<sup>9</sup>

Nous suivons la numérotation traditionnelle des classes. Les principaux allomorphes des préfixes de classe des nominaux indépendants:

4. Classe	Classe
1 m- (nas. syl.)	7 shi-
2 wa-	8 zi-
3 m- (nas. syl.)	9 IN-1 (nas. homorganique); 10-1
4 mi-	10 IN-1 (nas. homorganique); nyi-; 10-1
5 10°-1; d(i)-	11 u-
6 ma-	15 hu-, u-

Le préfixe de cl. 5 représenté par 10°-1 entraîne souvent le renforcement de l'initiale lexicale (passage de continue à occlusive); il peut être rapporté à un étymon bantou commun comportant une voyelle fermée (\*yi-). En effet, une voyelle de ce degré est, d'une façon générale, souvent reflétée en shingazija par le fait que la consonne suivante est une occlusive (voir Sibertin-Blanc 1980: 40 et Lafon 1987, vol. I: 259-65). La notation 10°-1 du préf. de cl. 5 permet de le distinguer de l'allomorphe 10-1 des cl. 9 et 10 que l'on trouve surtout dans les emprunts et qui représente une absence de préfixe sans modification de l'initiale lexématique.

Le préfixe nasal |N-1| des classes 9 et 10 s'amalgame avec l'initiale lexicale pour former un seul phonème prénasalisé |NC|, /~C/, qui comporte toujours, en principe, une occlusion: il participe de ce fait au conditionnement, en faisant correspondre à une continue une prénasalisée occlusive.

<sup>7</sup> Amalgame entre la nasale homorganique (jamais syllabique) et la consonne orale; m est utilisée devant les bilabiales, n dans les autres cas (nv, nd, etc.); le soulignement distingue les prénasalisées bilabiales des successions nasale-syllabique-bilabiale (mb [ ~ b ], mb [ m<sup>u</sup>b ], etc.).

<sup>8</sup> Après nasale, réalisation occlusive (mpv [ ~ b ], mpv [ m<sup>u</sup>b ], etc.).

<sup>9</sup> L'attaque forte est due à un hamza étymologique (ˀ), parfois à un ayn, alors que la nasalisation est toujours due à la présence d'un ayn étymologique.

Les préfixes verbaux sont toujours syllabiques, ils sont symbolisés par le préfix. de cl. 15, *hu-* ou *u-*, en variante partiellement libre (liée à la longueur de la racine). La forme de citation des verbes correspond à la forme minimale précédée d'un tiret qui indique la présence, nécessaire sauf à l'impératif, d'au moins un élément préfixé.

5. Rac: | -fung- | | notion d'attacher |
  - racine 'primaire':
    - verbe: -funga, attacher: | -fung-a |, rac. -suf. v.
    - nom: mfungo, noeud: | m-fung-o |, préf. cl. 3 - rac. -suf. n.
  - racine étendue: ext. sémantique: | -u- |, réversif:
    - verbe: -fungua, détacher: | -fung-u-a |, rac. -rév. -suf. v.
    - nom: mfunguo, clef: | m-fung-u-o |, préf. cl. 3 - rac. -rév. -suf. n.
  - racine étendue: ext. sém. synt.: | -w- |, passif:
    - verbe: -fungwa, être attaché: | -fung-w-a |, rac. - passif - suf. v.
    - racine étendue: ext. sém. + ext. sém. synt.: | -u- + -w- |:
    - verbe: -fungulwa, être détaché: | -fung-u-w-a |,
    - rac. - rév. - passif - suf. v. 10
6. Rac: | -l- | | notion de manger |
  - racine primaire:
    - verbe: -la, manger: | -l-a |, rac. - suf. v
    - nom: mlo, nourriture: | m-l-o |, préf. cl. 1 - rac. - suf. n
  - racine étendue: ext. synt.: | -is- |, causatif:
    - verbe: -lisa (...), nourrir (qqun): | -l-is-a |, rac. - caus. - suf. v
    - hulisa mwana, nourrir un enfant
  - racine étendue: ext. synt.: | -i- |, applicatif:
    - verbe: -lia (...), manger (avec): | -l-i-a |, rac. - appl. - suf. v.
    - hulia ha sutru, manger avec une cueillère

### 3.1.2 Constituants syntaxiques issus de thèmes

Les thèmes ne donnent que des nominaux et n'admettent pas de suffixe:

7. nominaux: préf. cl. + thème
8. Th: | -bibo | | notion de cajou |
  - bibo, fruit du cajoutier: | 0<sup>o</sup>-bibo |, préf. cl. 5 - thème
  - mbibo, cajoutier: | m-bibo |, préf. cl. 3 - thème
9. Th: | -ndru | | notion d'existant |
  - wandru, gens, personnes: | wa-ndru |, préf. cl. 2 - thème
  - hindru, chose: | hi-ndru |, préf. cl. 7 - thème
10. Th: | -wade | | notion de maladie |
  - mwade, malade: | mu-wade |, préf. cl. 1 - thème
  - uwade, maladie: | u-wade |, préf. cl. 11 - thème

Dans ces exemples, la reconnaissance de la base lexématique commune ne fait pas problème; dans ce cas, l'existence d'une relation entre ces mots est très généralement ressentie et acceptée par les locuteurs, ce qui justifie leur regroupement dans un même article du dictionnaire.

Mais ces exemples ne représentent pas l'éventail complet des possibilités lexicologiques de la langue: il convient de prendre en compte des cas où le

<sup>10</sup>Le -l- qui apparaît à la frontière morphologique est nécessaire car un glide apparaît entre toute succession vocalique (/ua/ est réalisé [uwa]); -l- peut donc être interprétée comme une consonne épenthétique ou comme appartenant étymologiquement au réversif, donné comme \*-ud- en bantu commun (c'est un rellet normal de \*d en intervocalique).

lexème ne se donne pas sous une forme rigoureusement identique dans les différents mots qui l'actualisent - ce qui oblige à fonder le regroupement sur une réflexion morphologique.

### 3.1.3 "Alternance consonantique"

Une alternance se produit entre continue et occlusive, le plus souvent à l'initiale du constituant syntactique, ce qui peut alors s'expliquer par un amalgame entre préfixe et initiale lexématique. La règle de distribution est que les continues apparaissent après préfixe syllabique (CV-, V-, m-), les occlusives après préf. 10<sup>e</sup>-1 de cl. 5 et 1N-1 de cl. 9/10. On trouve souvent en relation un verbe (initiale continue après préfixe syllabique) et un nominal en cl. 5 ou 9, ou encore, un nominal en classe 5 et le même en classe 6 (pref. 1ma-1), ces deux classes étant en opposition de nombre.

Dans cette analyse, la continue est la forme de base:

11. pv / p / mpb  
upvea, balayer |u-pve-a|, préf. cl. 15 - rac. - suf.v.  
→ peleo, balai |Ø-pve-e-o|, préf. cl. 5 - rac. - appli. - suf.n.<sup>11</sup>  
upvenya, cligner des yeux |u-pveny-a|, préf. cl. 15 - rac. - suf.v.  
→ mpbenye, éclair |N-pveny-e|, préf. cl. 9 - rac. - suf.n.  
mapvaha, chats |ma-pvaha|, préf. cl. 6 - thème  
→ paha, chat |0<sup>e</sup>-pvaha|, préf. cl. 5 - thème
12. w, 0 / b / mb  
hunda, construire |hu-wund-a|, préf. cl. 15 - rac. - suf.v.  
→ bundo |0<sup>e</sup>-wund-o|, préf. cl. 5 - rac. - suf.v.  
uwana, lutter |u-wan-a|, préf. cl. 15 - rac. - suf.v.  
→ mbwana, combat |N-wan-a|, préf. cl. 9 - rac. - suf.n.  
maive, pierres |ma-wel|, préf. cl. 6 - thème  
→ bwe, pierre |0<sup>e</sup>-wel|, préf. cl. 5 - thème
13. r / tr / ndr  
urongoa, parler |u-rongo-a|, préf. cl. 15 - rac. - suf.v.  
→ ndrongoo, affaire |N-rongo-o|, préf. cl. 9 - rac. - suf.n.  
maravu, jones |ma-ravu|, préf. cl. 6 - thème  
→ travu, joue |0<sup>e</sup>-ravu|, préf. cl. 5 - thème
14. l / d / nd  
ultima, cultiver |u-lim-a|, préf. cl. 15 - rac. - suf.v.  
→ dima, agriculture |0<sup>e</sup>-lim-a|, préf. cl. 5 - rac. - suf.n.  
ndlima, agriculture |N-lim-a|, préf. cl. 9 - rac. - suf.n.  
malaho, maisons |ma-laho|, préf. cl. 6 - thème  
→ daho, maison |0<sup>e</sup>-laho|, préf. cl. 5 - thème
15. h / k / nk  
uhoa, tousser |u-ho-a|, préf. cl. 15 - rac. - suf.v.  
→ nkoos, toux |N-ho-o|, préf. cl. 9 - rac. - suf.n.  
mahoko, grand-mères |ma-hoko|, préf. cl. 6 - thème  
→ koko, grand-mère |0<sup>e</sup>-hoko|, préf. cl. 5 - thème

Il arrive, dans quelques cas lexicalisés, en particulier à l'initiale de verbes, que l'opposition entre continue et occlusive soit significative:

<sup>11</sup> Sur l'insertion d'un -l, voir note 9.



16. upvaha 'peindre' ↔ upaha 'se maquiller'  
 huru 'lancer' ↔ hurta 'jeter'  
 ulipva 'payer' ↔ udipva 'riposter'  
 uhua 'éplucher' ↔ ukua '(se) gratter'<sup>12</sup>

La même alternance peut aussi apparaître en position interne.<sup>13</sup>

17. nkope 'cil' ↔ ikope 'orgelet'  
 mpéepvo 'éventail' ↔ upépo 'vent'  
 ntsihu 'jour' ↔ usiku 'journée'

**3.1.4 Absence de racine primaire**

Dans de nombreux cas, les racines n'apparaissent que sous des formes que l'on analyse comme étendues; l'appareillement entre les différents termes implique alors qu'on postule une matrice commune, qui peut être une forme primaire ou une forme complexe (non réductible en l'état de la documentation), que nous représentons par un verbal précédé de +, que nous ne traduisons pas:

18. + angama, racine primaire postulée de la comparaison entre:  
 -angamia, (se) détruire: applicatif | -angam-1-a |  
 -angamiza, détruire (qqch.): causatif | -angam-iz-a |
19. + -lekea, racine complexe (applicatif?) postulée  
 de la comparaison entre:  
 -lekelea, s'installer: applicatif | -leke-le-a |  
 -lekeza, réparer (qqch): causatif | -leke-za-a |

Parfois les deux phénomènes (alternance et absence d'attestation de la racine primaire) doivent être invoqués simultanément à l'appui du rapprochement - ce qui peut contribuer à le rendre plus discutable:

20. + -pvaha, racine primaire  
 postulée de la comparaison entre  
 -pvahiza, emmener (qqun en voiture): causatif | -pvah-iz-a |  
 -pashia, embarquer (dans un véhicule):  
 applicatif | -pah-1-a |<sup>14</sup>

<sup>12</sup>L'initiale occlusive de ces verbes peut s'expliquer en ce qu'il s'agirait de formes originellement réfléchies, ce que le sens paraît appuyer: l'étymon de l'infixe réfléchi (actuellement *-dji-*) est donné comme \**ji-*, c'est-à-dire avec une voyelle fermée — on retrouve la même situation que pour le préfixe de cl. 5; on aurait alors: \**pāk-* → *-pvaha*; \**ji-pāk-* → *-paha*.

Cette hypothèse est appuyée par des formes qui apparaissent comme des vestiges d'une telle évolution: *uona* voir mais *udjiuona* se voir (se considérer comme). Voir Lafon (1987: vol. I, 262 ff.). En tout état de cause, ce phénomène n'est plus opératoire en shingazidja, *-dji-* n'entraînant pas de modification de l'initiale verbale: *uua*, 'tuer' *udjiuua*, 'se tuer'. Ce même phénomène se trouve, de façon vestigieuse également, en swahili: *kuamba* 'dire' mais, avec "renforcement" de l'initiale, *kudjiigamba* se vanter (G. Philippson, communication personnelle).

<sup>13</sup>La forme non attendue (avec occlusive à l'intervocalique) est sans doute un emprunt à une langue bantu voisine, vraisemblablement au swahili.

<sup>14</sup>Palatalisation normale de *h* devant *-i*.

### 3.2 Emprunts arabes

#### 3.2.1 Relations lexicologiques par le biais des processus bantu

Les termes d'origine arabe constituent le tiers des mots du corpus de notre thèse, soit 1019 mots sur 3619, ou le quart des articles (505 sur 2288).

Ces emprunts ne s'insèrent que partiellement dans le modèle "bantu" de formation des mots: en particulier, il existe un état des verbes où ceux-ci n'admettent pas de suffixe verbal et ont en lieu et place une voyelle conditionnée.<sup>15</sup>

Toutefois, sous certaines conditions, les emprunts sont naturalisés et s'insèrent pleinement dans les processus bantu; on peut alors isoler des éléments lexicaux analogues aux deux catégories lexématiques bantu.

#### 3.2.2 Éléments analogues aux racines nomino-verbales

La présence d'une extension postposée à un terme d'emprunt permet de construire des verbes et des nominaux selon les procédés bantu.<sup>16</sup>

Sur le nominal *djimla* 'somme' (< ar. *jumla*, 'totalité') se place l'extension causative *-is-*, d'où la racine étendue *l-djimlis-l*, qui donne le verbal *-djimlisa*, 'totaliser' et le nominal *mdjimliso*, 'total'.

De même:

21. ar. *gasada* 'avoir l'intention de' > -kusudi 'avoir pour but'  
+ appl. -i- kusudio 'vocation' (cl. 5)  
-kusudia 'avoir pour but'

22. ar. *rakis* 'bon marché' > rahisi 'bon marché'  
+ caus. -ish- -rahisisha 'faciliter'

Il peut arriver que ne soit attestée en shingazidja qu'une forme étendue: par symétrie avec la situation des termes bantu, nous postulons comme entrée un verbe nu non attesté, non traduit, précédé de + :

23. ar. *hūdūr* 'visite' > + -hūduru  
+ appl. -i- mhūdurio 'rassemblement'  
+ caus. -ish- -hūdurisha 'inviter'

#### 3.2.3 Analogues aux thèmes nominaux

Le terme arabe *tājir* 'commerçant' donne:

24. tadjiri 'personne riche' | 0°-tadjiri |, préf. cl. 5 - thème  
matajiri 'personnes riches' | ma-tadjiri |, préf. cl. 6 - thème  
utajiri 'richesse' | u-tajiri |, préf. cl. 11 - thème

De même:

<sup>15</sup>Même chose en swahili; voir Lafon 1985.  
<sup>16</sup>Les références à l'arabe sont tirées de Wehr 1976; nous suivons sa transcription à l'exception de *ayn*, ici *ε*; les racines arabes sont données en majuscules; la non-correspondance des timbres vocaliques internes, dans le cas des verbes, entre l'emprunt shingazidja et le mot arabe proposé comme source s'explique en grande partie par le fait que les verbes shingazidja apparaissent le plus souvent construits sur l'inaccompli arabe: or, la forme de citation des dictionnaires arabes, utilisée ici, est celle de l'accompli, dont la vocalisation diffère de celle de l'inaccompli.

25. ar. kaid 'ruse' >  
 mkaidi 'têtu'  
 wakaidi 'têtu'  
 ukaidi 'obstination'  
 et avec alternance l / d / nd:  
 |m-kaidi|, préf. cl. 1 - thème  
 |wa-kaidi|, préf. cl. 2 - thème  
 |u-kaidi|, préf. cl. 11 - thème

26. ar. limûn 'citron' >  
 mlilmu 'citronnier'  
 mlilmu 'citronniers'  
 ndimn 'citron'  
 |m-lilmu|, préf. cl. 3 - thème  
 |m-lilmu|, préf. cl. 4 - thème  
 |N-lilmu|, préf. cl. 9 - thème

Les processus bantu sont très productifs, tant pour les termes indigènes que pour les emprunts — du moins pour ces derniers lorsqu'une extension est impliquée.

### 3.2.4 Relations lexicologiques par le biais de l'arabe

Les apparentements mis en évidence ci-dessus relèvent des processus de type bantu; ce sont des exemples caractéristiques d'intégration morphologique.

Mais les emprunts sont aussi apparentés à travers leur langue source, l'arabe — l'on sait que l'arabe est une langue dont le lexique est structuré autour de racines consonantiques, que divers processus, notamment des variations du schème vocalique, actualisent pour aboutir aux différents constituents syntaxiques.

Cela donnera les séries telles que les suivantes, où la relation entre les différents termes ne peut être établie par le biais de processus opératoires en shingazidja mais où il fait toutefois peu de doute que l'appareillement soit reconnu par la grande majorité des locuteurs:

27. racine	arabe	shingazidja
FKR	fakara 'penser'	-fikiri 'penser'
	fikra 'pensée'	fikra 'pensée'
ZIM	zulm 'injustice'	dhulma 'fraude'
	zallâm 'tyran'	dhwalimn 'exploiteur'
	zâlama 'faire du tort'	-dhulumu 'voler'
eJB	ejab 'merveille'	adjabu 'merveille'
	(*)istacajaba 'être étonné'	-staâdjabu 'être étonné'

Du fait du contexte socio-culturel des Comores, le processus d'emprunt à l'arabe est toujours actuel; la productivité s'exprime ici par le fait que, si existe en shingazidja un terme venant d'une racine arabe donnée, et que le besoin se fasse sentir d'un terme apparenté relevant d'une autre catégorie grammaticale, celui-ci pourra être emprunté directement à l'arabe, sans impliquer nécessairement les processus de formation de mots bantu.

### 4. Principe de constitution des articles de dictionnaire

Le choix de regrouper les mots issus d'un même lexème ayant été fait, comment l'appliquer au cas du shingazidja compte-tenu de ses spécificités?

La première décision concerne l'entrée, qui commande la place alphabétique de l'article (voir note 1). Contrairement notamment aux dictionnaires arabes, les entrées sont des mots de la langue, qu'ils soient effectivement attestés ou seulement postulés, et non des lexèmes: d'une part, les variations morpho-phonologiques se produisant à l'initiale rendent difficile la formalisation des lexèmes, et d'autre part ces lexèmes sont des

entités artificielles dont les locuteurs, dans la situation exclusivement orale du comorien, n'ont pas conscience (même s'ils perçoivent globalement le fonctionnement lexicologique de leur langue). Il nous a donc paru que fonder le classement du dictionnaire sur de tels éléments le rendrait tout à fait déconcertant.

Les mots constituant les entrées seront toutefois choisis pour leur proximité formelle avec la racine - ainsi on choisira plutôt des verbes (sans préfixe) que des nominaux, et parmi ces derniers, pour les emprunts du moins, des nominaux à préfixe |0|, etc (exemples ci-dessous).

Nous examinons maintenant comment sont constitués les articles de chaque origine.

#### 4.1 Article bantu

Le regroupement lexicématique des mots est guidé par deux critères:

- morphologique: que l'analyse fasse apparaître le même élément lexicématique ainsi que les morphèmes mis en jeu par les processus réguliers de formation des mots;
- sémantique: que ces termes manifestent entre eux un rapport sémantique.

L'étymon éventuel de Guthrie, qui peut contribuer à justifier le regroupement, est indiqué<sup>17</sup>. On distingue deux types d'article selon que l'on a affaire à une racine nomino verbal ou un thème nominal.

##### 4.1.1 Racine nomino-verbale

L'entrée pour une racine nomino-verbale sera organisée de la façon suivante:

- étymon éventuel de Guthrie
- verbe sans extension à la forme de citation
- nominaux
- extension, sous une formalisation immédiate, suivie des termes étendus, verbes puis nominaux<sup>18</sup>

28.	< BC *-dǝ-: manger	[étymon ]
	-la: manger	[forme de citation]
	mlo (mi-) 'nourriture'	[nominal]
	+ -i-: -lila: manger avec qqch;	
	ulio (ndio): table à trois pieds	[terme étendu]
	+ -is-: -lisa: nourrir	
	+ -w-: -lwa: être mordu	

<sup>17</sup>Le caractère trop partiel et parfois aléatoire des reconstructions lexicales de Guthrie, ainsi que leur statut explicite de reconstructions synchroniques, ne saurait leur donner à notre avis une valeur déterminante quant à la décision d'appartenance.

<sup>18</sup>Nous ne donnons du contenu des articles que ce qui apparaît nécessaire à la démonstration; la formation du pluriel des nominaux est indiquée par le préfixe de classe donné entre parenthèses suivi d'un tiret; lorsque les préfixes pluriel et singulier sont identiques (cl. 9/10), seul figure un tiret; en cas de modification de l'initiale lexicale (alternance), la forme du pluriel figure en toute lettre; cette formalisation, classique, suffit à expliciter, sans ambiguïté, l'appartenance de classe de chaque nom.

Dans le cas de racine à alternance initiale, on commence par la forme jugée régulière (à initiale continue):

29. < BC \*-pàk-, froter [étymon]  
 -pvaha: peindre [forme régulière]  
 mpvaho (mi-): peinture [forme "renforcée"]  
 -paha: se maquiller  
 mpaho (mi-): maquillage

Dans le cas où seules des formes étendues sont attestées, l'entrée est constituée du verbe issu de la racine hypothétique reconstruite (qui peut être complexe):

30. + -lekea [racine hypothétique]  
 + -le-: -lekelea: s'installer [termes étendus]  
 + -z-: -lekeza: installer

#### 4.1.2 Thème nominal

Le nominal jugé le plus neutre sémantiquement détermine l'entrée, suivi des formes marquées (augmentatifs, diminutifs, etc):

31. < BC \*-búdi, chèvre [étymon]  
 mbuzi (-): chèvre [forme neutre]  
 buzi (ma-): bouc; grosse chèvre [forme marquée]
32. < BC \*-púko, sac [étymon]  
 mfuko (mi-): sac, poche [forme neutre]  
 fuko (ma-): chambre [formes marquées]  
 ifuko (zi-): petite chambre

Certains choix arbitraires ont été faits, comme de commencer systématiquement par le nom de la plante plutôt que par celui de l'arbre (obtenu par la mise en classe 3/4, préf. m-/mi-):

33. nkungu-manga (-): noix de muscade [plante]  
 nkungu-manga (mi-): muscadier [arbre]

#### 4.2 Article d'emprunt à l'arabe

Le regroupement est fondé sur un double critère:

- la reconnaissance des processus morphologiques de type bantou — on suivra alors la même présentation que pour les articles bantou;
- la lexicologie arabe: tous les termes shingazidia provenant de mots arabes issus eux-mêmes d'une racine reconnue comme une dans la lexicographie arabe sont regroupés, quelle que soit leur forme.<sup>19</sup> Le recours à l'étymologie apparaît en effet comme le seul moyen non arbitraire d'aboutir, pour les emprunts à l'arabe, à un regroupement lexicographique comparable à celui des termes bantou, tout en prenant en compte la productivité, c'est-à-dire la possibilité d'emprunt de nouveaux termes sur des racines déjà représentées, et de rendre par là même l'organisation du dictionnaire cohérente.

La racine arabe, qui justifie le regroupement des termes, est donnée, ainsi que, pour chaque terme, le cas échéant, la source précise.

<sup>19</sup>Notre référence étant le dictionnaire de Wehr 1976.

34. < ar BDL  
 -badlii: changer [racine arabe]  
 [forme de citation]  
 [source]  
 + -ish-: -badliisha: changer qqch; [forme de citation]  
 mbadliishadij (ma-): réformateur [termes étendus]  
 badala (-): équivalent [source]  
 [ar. badal, substitut]
35. < ar. FSR  
 -fasiri: expliquer; [racine arabe]  
 [forme de citation]  
 [nominal]  
 [source]  
 fasiri (-): explication [source]  
 [ar. fassara, expliquer]  
 -tafasiri: même sens que -fasiri; [termes dérivés par l'arabe]  
 tafsiiri (-): traduction [source]  
 [ar. tafsiir, explication] [source]
5. Problèmes découlant du mode de constitution des articles  
 On perçoit les difficultés que ce type de choix lexicographique entraîne; elles se présentent différemment selon l'origine des termes.
- 5.1 Emprunts
- Le regroupement des termes à travers la lexicographie arabe permet de justifier des rapprochements évidents, tels ceux mentionnés plus haut, où cependant aucun procédé de dérivation bantu n'est en jeu. L'inconvénient est que certains apparentements sont susceptibles de paraître forcés à tout locuteur comorien et peut-être aussi, d'ailleurs, à la plupart des locuteurs arabes non familiarisés avec la tradition lexicographique de leur langue. C'est sans nul doute le cas des termes shingazidja *-dai* et *dua*, issus tous deux de la racine arabe *DEW*, et donc rangés dans un même article:
36. ar. DEW →: [racine arabe]  
 ar. (i)ddacä, réclamer > -dai, exiger, [sources et termes]  
 [shingazidja]  
 ar. ducaä, supplique > dua, bénédiction [sources et termes]  
 [shingazidja]
37. ar. KLF: [racine arabe]  
 De même pour *halafu* et *hitlafi*, de la racine *KLF*:  
 ar. kalif, derrière > halafu, ensuite, [sources et termes]  
 [shingazidja]  
 ar. ikliläf, désaccord > hitlafi, divergence [sources et termes]  
 [shingazidja]
- Ainsi que pour *ada* et *ndi* (€WD), *dari* et *diwara* (DWR):
38. ar. €WD: [racine arabe]  
 ar. äda, habitude > ada, grand mariage [sources et termes]  
 [shingazidja]  
 ar. eüdü, luth > udi, luth [sources et termes]  
 [shingazidja]
39. ar. DWR: [racine arabe]  
 ar. dâr, maison > dari, étage [sources et termes]  
 [shingazidja]  
 ar. dawwâr, en rotation > diwara, roue [sources et termes]  
 [shingazidja]

On pourrait certes envisager dans de tels cas de ne pas suivre strictement le principe de regroupement étymologique et de dissocier les termes shingazidja issus de la même racine arabe dont il est clair que la parenté, sur les plans formel et sémantique, n'est jamais ressentie. Cette idée, évidente en apparence, se révèle d'une application difficile: comment distinguer les rapprochements ressentis de ceux qui ne le sont pas?

## 5.2 Termes bantou

Il semblerait que le regroupement lexématique des termes bantou, s'appuyant sur des processus morphologiques identifiés et opératoires en shingazidja, ne devrait guère poser de problème; or, dans un nombre non négligeable de cas, ces critères ne sont pas suffisants.

La discussion est ici d'ordre sémantique: à côté en effet des cas simples et peu discutables, même lorsqu'est impliquée une alternance consonantique ou la reconstruction d'une forme primaire, il en existe d'autres où la relation sémantique éventuelle entre les formes considérées n'est pas évidente. Rappelons que la profondeur historique, dans les langues à "tradition orale", fait défaut et que l'on ne peut donc se référer à un état antérieur de la langue qui expliquerait des évolutions sémantiques... Le comparatisme peut fournir des éléments mais, en l'état de la documentation, ils sont rarement suffisants. On est donc amené finalement à faire reposer la décision sur l'arbitraire ou l'intuition du lexicographe.

Nous donnons quelques exemples de rapprochements sémantiquement discutables (certains ont été opérés dans le lexique, d'autres, non):

40.	-umba	'créer'
	nyumba (-)	'maison'
	-baya	'fermer'
	-bua	'ouvrir' (en supposant un réversif -u-)
	-hoha	'griller'
	djosi (cl. 5)	'fumée' (en supposant la chute du h initial, ce qui est fréquent)
	-katra	'couper'
	mkatre (mi-)	'galette', tranche de pain (ce qui est coupé)

Ces difficultés toutefois ne nous paraissent pas devoir mettre en cause la validité du principe d'organisation par article lexématique: elles ne concernent qu'un nombre limité d'articles; par ailleurs, l'intégration dans le corps de l'ouvrage de renvois pour tous les termes n'apparaissant pas à leur place alphabétique — même lorsqu'il s'agit de formations régulières — permet de retrouver tout terme, quelle que soit la place où il a été assigné.

## Conclusion

On voit que la tâche du lexicographe est ici loin de se réduire à la simple compilation des données relevées; le lexique procède en fait d'une analyse morphologique de la langue, sur laquelle il se fonde et qu'il fonde en même temps. On rejoint là la définition que le professeur Houis (1983: 9) donnait de la lexicologie systématique, qui

...relient comme objet la systématique de la langue dans la mesure où il s'agit d'identifier des unités comme le mot. Elle fournit par là même un fondement rigoureux à l'organisation lexicographique.

Et c'est cela qui fait selon nous l'intérêt du travail lexicographique, tout en le rendant d'autant plus vulnérable au passage du temps: le découpage lexicographique proposé ne saurait être fixé d'une manière définitive, il suit les avancées dans la connaissance de la langue. La découverte d'un chaînon lexicographique manquant, qui peut provenir d'un affinement d'une définition ou de la comparaison avec les langues voisines, est en effet susceptible d'aboutir soit au regroupement de termes jusqu'alors dissociés, soit au contraire à l'éclatement d'articles...

Centre d'Etudes sur l'Océan Indien (CERO)  
INALCO  
2, rue de Lille, Paris VII, France

#### Références

- Ahmed-Chamanga M. et N.-J. Guenier 1977, Recherches sur l'instrumentalisation du comorien, *Cahiers d'Etudes Africaines*, 66-67: vii: 213-39).
- 1979 *Le dictionnaire Comorien-français et Français-comorien du R.P. Saclaux*, Paris: Selaï.
- Ahmed-Chamanga M., M. Lafon et J.-L. Sibertin-Blanc 1988. Projet d'orthographe pratique du comorien, *Etudes Océan Indien* 9: 7-35.
- Alexandre, P. 1981 Les langues bantu, dans *Les langues dans le monde ancien et moderne* (Paris: CNRS) p. 353-75.
- Benson 1964 A century of Bantu lexicography, *African Language Studies* 5: 64-91.
- Blanchy, S. 1987 *L'interprète mahorais-français et français-mahorais*. Mayotte.
- Chanudet, Cl. 1988 *Contribution à l'étude de l'origine du peuplement de l'île de Mohéli*, thèse de 3<sup>e</sup> cycle, Paris-INALCO.
- Fisher, F. 1949 *Grammaire-dictionnaire comorien*. Strasbourg: éditions de la Basse-Alsace.
- Guthrie, M. 1967 *Comparative Bantu*. London.
- Hill, T. 1973 The primary dialects of Swahili: an approach to a linguistic geographical survey, *Swahili* (Dar-es-Salaam) 43.2: 7-14.
- Houis, M. 1983 Des unités significatives, préalable à la lexicographie, *Afrique et Langage* 19: 5-31.
- Lafon, M. 1985 Régularité et irrégularité dans le système verbal du shingazidja — la voyelle finale des thèmes verbaux, *Afrique et Langage* 22: 5-33.
- 1987 *Le shingazidja, une langue bantu sous influence arabe*, Paris-INALCO.
- 1989 Situation linguistique à la Grande-Comore, essai de définition du statut de l'arabe, *Matériaux Arabes et Sud-arabiques*.
- Lafon, M. et J.-L. Sibertin-Blanc circa 1976 Propositions pour une graphie du comorien (n.d.).
- Sibertin-Blanc, J.-L. 1980 Sur quelques aspects des dialectes comoriens en contraste avec le kiswahili, *Etudes et documents* (Zanzibar) 1: 33-68.
- Saclaux, Ch. 1939 *Dictionnaire Swahili-Français*. Paris: Institut d'Ethnologie.
- Wehr 1976 *Arabic-English Dictionary*. Ithaca: Cornell University Press.



