

## **Vers un dictionnaire orthographique bambara**

*Valentin Vydrin*

*INALCO — LLACAN, Paris*

*Université d'État de St. Petersbourg*

*vydrine@gmail.com*

### **1. Les préalables**

De nombreux efforts pour standardiser le mandingue et le bambara en particulier ont produit certains résultats : les règles de base ont été formulées. Notre article précédent sur l'orthographe bambara (Konta & Vydrin 2014) a été une tentative d'avancer davantage dans la même optique pour couvrir des zones qui étaient restées grises : l'écriture des séquences du type *Nin* et *Nun* ; la notation tonale, etc.

Depuis la parution de cet article, il y a eu un progrès considérable en ce qui concerne les bases de l'étude du bambara écrit. Le facteur déterminant a été la croissance et la modernisation du Corpus Bambara de Référence (Vydrin et al. 2011), ce qui a facilité une étude statistique des occurrences des variantes orthographiques des lexèmes bambara dans les textes écrits, publiés sur papier ou affichés dans l'Internet. L'existence d'un tel outil a rendu faisable le projet d'un dictionnaire orthographique bambara basé sur le Corpus Bambara de Référence. Dans ce dictionnaire, le choix entre les variantes graphiques d'un lexème ferait en fonction, avant tout, de la fréquence des variantes dans les textes.

Le concept d'un tel dictionnaire a été formulée dans l'article (Vydrin & Méric 2020). Je vais en reproduire les idées principales ici, sans trop entrer dans les détails.

On s'est basé sur la version en cours du dictionnaire Bamadaba (Bailleul et al. 2011). Parmi les plus de 13 365 lexèmes de ce dictionnaire,<sup>1</sup> presque 4000 avaient des variantes graphiques. On a extrait de Bamadaba ces lexèmes pour lesquels il fallait choisir une variante standard.

Puis, Jean-Jacques Méric a écrit un script permettant d'obtenir automatiquement les données statistiques suivantes pour chaque variante de chaque lexème :

---

<sup>1</sup> Au moment de l'étude statistique. Le dictionnaire Bamadaba évolue en permanence ; au moment où j'écris cet article, le nombre de lexème dans Bamadaba a atteint 13 500.

- le nombre d’occurrences dans le sous-corpus désambiguïsé et sous-corpus non-désambiguïsé ;
- le nombre de textes dans le sous-corpus désambiguïsé et sous-corpus non-désambiguïsé où la variante apparaît (et aussi, les titres de ces textes) ;
- le nombre des auteurs utilisant cette variante, dans le sous-corpus désambiguïsé et non-désambiguïsé.

Les statistiques concernant le nombre des textes et des auteurs sont indispensables pour contourner l’ « effet de buccin » (Kilgarriff 1997: 138–139). Cet effet consiste en ce que la fréquence d’un mot (qui est globalement assez rare dans les textes) peut être anormalement augmentée si on inclut dans le corpus un document s’agissant de l’objet ou du phénomène en question (par exemple, une étude zoologique portant sur le buccin). Dans notre cas, une variante graphique est vue comme plus répandue si elle apparaît dans un plus grand nombre de textes différents et est utilisée par un plus grand nombre d’auteurs ; on peut même donner la préférence à ces facteurs par rapport au nombre global des occurrences des variantes dans le Corpus.

Les résultats de l’application du script ont été représentés sous format d’un tableau. Ci-dessous, je présente un extrait de ce tableau où les statistiques sont données concernant le lexème qui a quatre variantes graphiques : *fàsa*, *pàsa*, *fàsan*, *pàsan*, et fonctionne comme verbe avec les sens suivants : ‘maigrir, faire maigrir, rendre consistant, rendre coriace, s’obstiner à nier’.

variante	# occ simple	# occ désamb.	# sources	# sources	# auteurs	auteurs
fàsa	872	157	99	dogotoro_11balo_nafama.dis.html (13), tarawele-baganlatolola.dis.html (5), dumestre-chroniques_amoureuses_1995_11_14.dis.html (5), gorog_meyer-contes_bambara1974_16.dis.html (4), dumestre-manigances_2003_03.dis.html (4)	37	UNDEF (42), Touré, Seydou (27), Chan, Michelle & Heberger, Thomas & Werner, David (21), Tarawele, Daramani (6), Jara, Døkala Yusufu (5)
pàsa	128	15	11	saheli05_04danbele-baganmaralaw_ka.repl.html (4), kibaru110_03koroma-gundo.old.repl.html (2), sunbunu-fula_npogotigi-zup.old.repl.html (1), sankore08_09jigifa-banyereye-zup.old.repl.html (1), kibaru115_08tobilike_dibi.old.repl.html (1)	6	UNDEF (4), Danbele, Alu (4), Koroma, Bayidi (2), Ture, Basoso (Basaro) (1), Touré, Seydou (1)
fàsan	96	7	1			
pàsan	35	1	1			

Tableau 1. Statistique pour le lexème *fàsa* ‘maigrir’

Pour ce lexème, toutes les données statistiques convergent : par tous les paramètres, la variante *fàsa* dépasse de loin les trois autres variantes, il doit donc être choisi comme variante standard, et les autres sont à déconseiller.

En mai-juin 2020, nous (Jean-Jacques Méric et moi) avons effectué une première sélection. Pour les environ quatre mille lexèmes ayant des variantes graphiques, nous avons essayé d'identifier les variantes qu'on pourrait recommander comme standards. Plus précisément, nous avons cherché à trier les variantes en trois types :

- « r » : variante recommandée comme standard ;
- « t » : variante tolérée ;
- « e » : variante déconseillée (à rejeter).

Pour chaque lexème, une seule variante peut être recommandée. On marque comme « tolérées » les variantes dont la fréquence est inférieure à celle de la variante « recommandée », mais demeure cependant assez élevée.<sup>2</sup> On cherchait à éviter la prolifération des variantes « tolérées » (car le but d'un dictionnaire orthographique est la normalisation de la langue écrite, donc la minimisation de la variation), cependant, elles sont assez nombreuses dans notre liste, et un bon nombre de lexèmes a deux (ou même plus de deux) variantes « tolérées ».

Toutes les autres variantes graphiques sont « rejetées ».

Cependant, une approche purement statistique s'est avérée insuffisante. Il y a eu des lexèmes pour lesquels les données statistiques ont été trop faibles (pas ou très peu d'occurrences dans le Corpus), ou, d'autre part, pour lesquels les statistiques d'occurrences semblaient être en contradiction avec notre intuition. Tous les lexèmes pour lesquels nous avons eu des doutes ont été mis à part ; leur nombre a légèrement dépassé 1200.

A l'étape suivante, la liste de ces 1200 lexèmes a été débattue par un groupe d'experts dans le cadre d'un atelier de travail organisé par l'AMALAN (l'Académie Malienne des Langues Nationales) du 30 mai au 18 juin 2022 à Bamako (cet atelier a bénéficié d'une subvention de Lacuna Fund (une initiative du Rockefeller Foundation, Google.org et Canada's International Development Research Centre). Voici la composition du groupe d'experts et leurs affiliations :

AMALAN : Dr. Adama Diokolo Coulibaly (le Directeur), Mme Touré Aoua Samake, Dr. Mahamadou Konta, Adama Dembele.

---

<sup>2</sup> Il n'a pas été facile d'établir des critères numériques strictes, d'autant plus que d'autres facteurs ont été pris en compte aussi. Juste pour donner une idée : si la fréquence de la première variante était 2 ou 3 fois plus élevée que celle de la 2<sup>e</sup> variante, nous marquons cette 2<sup>e</sup> variante comme « tolérée ». Et si le dépassement était de l'ordre de 10 fois, on marquait la 2<sup>e</sup> variante comme « rejetée ».

École Normale Supérieure : Dr. Ibrahima Diawara.

Université des Lettres et des Sciences Humaines de Bamako (Université de Kabala) : Dr. Issiaka Ballo.

DNENF-LN (Direction Nationale de l'Éducation Non-formelle et Langues Nationales) : Diakaridia Diarra.

CNR-ENF (Centre National des Ressources de l'Éducation Non-Formelle) : Joseph Drabo.

DNP (Direction Nationale de la Pédagogie) : Mme Ndiaye Fatoumata Keïta.

DNEF (Direction Nationale de l'Enseignement Fondamental) : Mme Salimata Coulibaly.

Kibaru (journal) : Youssouf Dokala Diarra.

CALAN-Mali (Club des Amis des Langues Nationales du Mali) : Mme Kadidia Toure.

Robots-Mali : Mouktar Traoré.

INALCO-LLACAN, France : Valentin Vydrin.

Il faut souligner que le grand souci des organisateurs de l'Atelier avait été d'assurer la participation des instances maliennes impliquées dans la politique linguistique ; par contre, on n'avait pas l'ambition d'assurer la représentativité de toutes les zones dialectales principales. Ainsi, les dialectes de Bélé Dougou et de Baninko ont été bien représentés, tandis que certaines autres zones n'ont pas été couvertes.

Dans ce qui suit, je présenterai les résultats du travail de cet atelier.

## 2. Règles d'orthographe revisitées

Avant d'examiner la liste de 1200 lexèmes, il a été décidé de faire d'abord un tour des règles d'orthographe bambara déjà établies. Il s'agit, avant tout, des règles formulées dans les publications (Anonyme 1979; Anonyme 1993), mais aussi dans l'article (Konta & Vydrin 2014) qui n'avaient pas encore été sujet de discussion parmi les experts.

### 2.1. Les règles des « Guides d'orthographe » qui ont été confirmées sans discussion

#### 2.1.1. Consonne vélaire intervocalique

Dans la position intervocalique, à l'intérieur d'un pied,<sup>3</sup> la consonne vélaire s'écrit *g* : *tɔ́gɔ́*<sup>4</sup> 'prénom' (et non pas *tɔ́kɔ́*, ni *twáa*). Les exceptions : *táa* (la variante

---

<sup>3</sup> Sur le pied caractéristique en bambara, voir (Vydrin 2020a).

<sup>4</sup> Dans l'orthographe malienne du bambara, les tons ne sont pas marqués. Cependant, dans cet article, je le notes, selon le modèle du Corpus Bambara de Référence.

la plus usitée même à Bamako, à côté de *tága*) ; *màa* ‘homme, personne’ comme une variante facultative (« tolérée »), à côté de *mògò*.

### 2.1.2. Consonne vélaire intervocalique précédée d’une nasale

Là où on observe à l’intérieur du mot une variation dialectale entre *ng* ~ *nw* ~ *ŋ*, on retient les formes avec *-ng-*. Ainsi, parmi les variantes *fànga* ~ *fànwa* ~ *fàŋa* ‘pouvoir’, on retient *fànga* ; entre *dìnge* ~ *dìŋe* ‘trou’, on retient *dìnge*.

### 2.1.3. Élision vocalique entre les consonnes

Là où la voyelle entre deux consonnes tend à se réduire jusqu’à l’élisio

### 2.1.4. Consonnes prénasalisées à l’initiale

Ces consonnes manifestent souvent une variabilité : la consonne sourde dans de nombreux dialectes correspond à une consonne sonore dans d’autres, par exemple :

	‘cache-sexe des filles’	‘mensonge’	‘crapaud’	‘baobab’
bamanankan standard	<i>npògo</i>	<i>nkàlon</i>	<i>ntòri</i>	<i>nsira</i>
bamanankan de Ségou	<i>nbògo</i>	<i>ngàlon</i>	<i>ndòri</i>	<i>nzira</i>

Dans les cas pareils, il est recommandé d’écrire la consonne sourde, comme en bamanankan standard : *npogo*, *nkalon*, *ntori*, *nsira*.

### 2.1.5. Les constructions nominales déterminatives Nom + Nom

Ces constructions, qui sont tonalement compactes (possédant un seul contour tonal), s’écrivent collées, quel que soit leur degré de figement :

*mìsisen* [mìsisén] ‘patte de bœuf’, *màloko* [màlòkó] ‘affaire de riz’, *mògòbajekulu* [mògòbàjékúlú] ‘association des adultes’.

Par contre, les constructions génitives (inaliénables) Nom + Nom, tonalement non-compactes, s’écrivent séparées :

*jèkulu` jémaa`* [jèkúlú` jémáa`] ‘dirigeant de l’association’, *mùsomannin` sì`* [mùsòmànnín` sǐ`] ‘l’âge de la jeune fille’.

### 2.1.6. Constructions attributives

Les constructions attributives (Nom + Adjectif ou Nom + Participe) s’écrivent séparément, même celles qui sont tonalement compactes :

<sup>5</sup> Voir (Vydrin 2020b) et (Vydrin 2020a) sur le conditionnement de l’élisio

Nom + Adjectif, tonalement compactes : *sò bilen* [sòbílén] ‘cheval rouge’, *wùlu juguman* [wùlùjúgúmán] ‘le chien méchant (parmi tous les chiens)’, *síra kura* [sírákúrá] ‘nouvelle route’. Seules les formes lexicalisées sont écrites collées, ex. : *néjugu* ‘parasite’, *dáje* ‘antilope koba’, *bàgifin* ‘acacia spirale’, *báncinin* ‘soeur cadette de la mère’, etc.

Nom + Adjectif, tonalement non-compactes : *ɲɔ̀ bùgurima* [ɲɔ̀ bùgùrimá] ‘mil poussiéreux’, *dúnán sènnateli* [dúnán sènnàtélí] ‘étranger pressé’.

Nom + Participe, tonalement non-compactes : *mùso sìgilen* [mùsó sìgílén] ‘une femme assise’, *dógó fèereta* [dógó fèeretà] ‘le bois pour la vente’, *dén sùnògòbali* [dén sùnògòbáli] ‘l’enfant qui n’a pas dormi’.

### 2.1.7. Les constructions distributives

Les constructions distributives à connecteur *ô* s’écrivent séparément :

*dùgu ô dùgu* ‘n’importe quel village’, *sògo ô sògo* ‘quelle que soit la viande’.

## 2.2. Des propositions formulées dans (Konta & Vydrin 2014)

### 2.2.1. Séquences « *f*-, *s*- suivies de voyelles semi-ouvertes »

Les voyelles moyennes longues provoquent une palatalisation facultative des consonnes *s*, *f*; dans ce cas, la voyelle peut se réaliser comme brève : *séere* ~ *shyéere* ‘témoin’, *fèere* ~ *fyèere* ‘vendre’, *fèere* ~ *fyèere* ‘ruse’. Après *s*, la même variabilité concerne la voyelle *ɔɔ* : *sɔɔ́lɔ́* ~ *shyɔ́lɔ́* ‘égoutter’, *sɔɔ́ri* ~ *syɔ́ri* ‘insérer’.

La décision : ***On retient la variante à voyelle longue, sans palatalisation : seere, feere, fèere, sɔɔ́lɔ́, sɔɔ́ri.***

### 2.2.2. Syllabes « consonne nasale + voyelle *i* ou *u* »

Dans les syllabes du type « consonne nasale + voyelle *i* ou *u* » à la fin du mot le caractère nasal ou oral de la voyelle n’est pas pertinent : *ɲún* [ɲú ~ ɲũ] ‘charger (sur la tête)’. Suivant les résultats d’une étude statistique, il est proposé de transcrire les syllabes de ce type par défaut sans *-n* final (*ɲíni* ‘chercher’, *mù* ‘enduire’, *dími* ‘douleur’), sauf dans les mots et morphèmes suivants où la présence de *-n* final prédomine dans les textes et peut être considérée comme consacrée par la tradition (cela concerne également les mots composés et dérivés comportant ces racines) :<sup>6</sup>

*dùnun* ‘tambour’, *fúnun* ‘gonfler’, *kínin* ‘droite’, *kúnun* ‘(se) réveiller’, *kùnun* ‘avalier’, *kúnùn* ‘hier’, *mìn* ‘boire’, *mîn* marque relative, *mín* ‘où ?’, *mùn* ‘quoi ?’, *nìn*

<sup>6</sup> Dans (Konta & Vydrin 2014) il a été proposé d’écrire le mot pour ‘partisan, allié’ avec un « n » à la fin (*ɲin*), ce qui le distinguerait du verbe qualitatif *ɲi* ‘être bon’ (les deux lexèmes étant étymologiquement liés). Cependant, les participants de l’Atelier se sont prononcés pour la forme *ɲi* pour les deux lexèmes.

‘ce’, *-nin* suffixe diminutif, *nún* ‘nez’, *nùn* ‘grâce à’, *ɲín* ‘dent’, *ɲún* ‘charger (sur la tête)’, *túnun* ‘disparaître’.

### 2.2.3. Lettres z et v

La lettre *z* ne s’écrit que dans des emprunts (surtout au français), ex. : *ziɲɔri* ‘junior’, *zido* ‘judo’, *zuwɛn* ‘juin’. Pour les mots d’origine mandingue où *z* apparaît comme une variante, à côté des consonnes prénasalisées *ns* et/ou *nz*, on retient la variante *ns*, ex. : *nson* ‘voleur’ (plutôt que \**nzon* ou \**zon*).

La lettre *v* n’est pas utilisée en bamanankan ; dans les mots d’emprunt, y compris les noms propres, on écrit un *w* : *wɛri* ‘verre’, *Waleri* ‘Valérie’.

### 2.2.4. L’orthographe de /ʃ/

Selon les règles de l’ancienne orthographe du bamanankan, on transcrivait /ʃ/ par le digraphe *sh* (Anonyme 1979: 11). Après la réforme orthographique des années 1980, il a été décidé de ne pas distinguer /ʃ/ de /s/ (Anonyme 1993: 10) ; on prescrit l’orthographe *sɔ* pour /ʃɔ/ ‘haricot’ et *sɛ* pour /ʃɛ/ ‘poulet’. La raison en a été la suivante : dans la plupart de dialectes du bamanankan, le phonème /s/ a un allophone [ʃ] dans des contextes qui varient d’un dialecte à l’autre, et dans la perspective pan-dialectale, il s’agit de la variation libre : [bási ~ bàʃi] ‘couscous’, [sú ~ ʃú] ‘nuit’, etc.

Cependant, on trouve en bamanankan au moins quatre mots où la variation entre *s* et *ʃ* n’existe pas, et on prononce invariablement un *ʃ* : *ʃɛ* ‘poulet’, *ʃɔ* ‘haricot’, *ʃù* ‘chou’, *ʃán* ‘herbe *Rottboellia exaltata*’ (dans le dialecte Bélé Dougou, *ʃɛn*). Pour deux de ces mots, on trouve même des paires minimales :

*ʃù* ‘chou’ – *sù* ‘cadavre’,

*ʃán* ‘herbe *Rottboellia exaltata*’ – *sán* ‘an, année’.

Pour ces quatre mots, il a été décidé de recommander l’utilisation du digraphe *sh*, mais, pour la continuité, de tolérer également l’écriture avec un *s*. On a donc les variantes graphiques :

*she* ~ *sɛ* ‘poulet’

*shɔ* ~ *sɔ* ‘haricot’

*shu* ~ *su* ‘chou’

*shan* ~ *san* ‘herbe *Rottboellia exaltata*’.

Tous les autres lexèmes s’écrivent avec un *s* seulement : *si* ‘âge’, *su* ‘nuit’, *basi* ‘couscous’, etc.

### 2.2.5. La voyelle nasale suivie d’une consonne nasale

La nasalité vocalique est désignée par la lettre *n* suivant la voyelle. Pourtant, dans les publications, là où une voyelle vocalique apparaît devant une consonne nasale (surtout en composition ou en dérivation), on omet souvent la lettre *n* désignant la nasalité vocalique.

Les experts ont confirmé que l'écriture de *n* dans de pareils contextes est obligatoire, son omission est inadmissible. Par exemple :

Écriture correcte	Écriture incorrecte	Traduction
<i>denmisennin</i>	* <i>demisenin</i>	'petit enfant'
<i>bagenna</i>	* <i>bagena</i>	'chevrier'
<i>A siranna.</i>	* <i>A sirana</i>	'Il a eu peur'.

Il en est de même pour les voyelles nasales suivies des suffixes *-ntan* (privatif) et *-nci* (marque de l'agent excessif). Dans ce contexte, une double nasale s'écrit :

Écriture correcte	Écriture incorrecte	Traduction
<i>senntan</i>	* <i>sentan</i>	'qui n'a pas de jambes'
<i>kunntan</i>	* <i>kuntan</i>	'sans tête ; inutile'
<i>nkalonnci</i>	* <i>nkalonci</i>	'menteur invétéré'

Sinon, on confondrait (surtout en écriture non-tonalisée) *sen-ntan* 'qui n'a pas de jambes' et *se-ntan* 'personne sans pouvoir' ; *kun-ntan* 'sans tête' et *ku-ntan* 'qui n'a pas de queue', etc.

La double ou même triple nasale s'écrit dans des mots conglomérés où cela est nécessaire pour représenter la structure morphologique de ces mots, comme *bìnnkanni* 'agression, viol' (plutôt que \**bìnkani*) ou *j̀̀ǹ̀ná* 'mépris' (plutôt que \**j̀̀nna*).

### 2.2.6. Postpositions

Les postpositions s'écrivent séparément des groupes nominaux précédents dont elles sont tonalement autonomes.<sup>7</sup>

De point de vue de leurs composition, il y a des postpositions simples et composées.

Voici la liste exhaustive des postpositions simples :

*bára* 'chez',

*bólo* (valeur de possession et de contrôle),

*cé* 'entre',

*fê* 'vers',

*kàn* 'sur',

*kó* 'derrière, après',

*kóro* 'sous',

*kóno* 'à l'intérieur de',

*kùn* 'sur la tête de, en possession de',

<sup>7</sup> Sauf les cas où elles entrent dans des mots composés, tonalement compacts, ex. : *kófesen* 'patte de derrière'.



*lá/ná* (valeur locative générale),  
*mà* (valeur adressative),  
*ɲé* ‘devant’,  
*yé* (valeurs équative, adressative, de transformation etc.),  
*yóro* ‘chez’.

Les postpositions composées comportent deux parties : une partie nominale (*kùn-* ‘tête’, *kèrɛ-* ‘côte’, *ɲé-* ‘œil, devant’, etc.) suivie d’une postposition simple. Ces composantes ne sont pas tonalement compactes, chacune garde son ton original. Cependant, on les considère comme des mots composés, plutôt que des combinaisons de deux mots (nom + postposition), par le critère d’impossibilité d’insertion d’autres mots. En bamanankan, un mot passe-partout est la particule de focalisation *dè*, on l’utilise pour vérifier la séparabilité. Cette particule ne peut pas être insérée entre les composantes d’une postposition composée. Ainsi, on peut dire :

(1a) *Û be tɛmɛ ɲógɔn dè kèrɛfɛ.*  
 3PL IPFV passer RECP FOC par.côté  
 ‘Ils passent à côté l’un de l’autre’.

Par contre, on ne peut pas dire :

(1b) \**Û be tɛmɛ ɲógɔn kèrɛ dè fɛ.*  
 3PL IPFV passer RECP côté FOC par

Voici la liste exhaustive des postpositions composées :

*bólokɔro* ‘sous la main de’,  
*céla* ‘au milieu de’,  
*céfɛ* ‘au milieu de’,  
*cémà* ‘au milieu de’,  
*dáfɛ* ‘près de’,  
*dála* ‘à côté de’,  
*jùlá* ‘au fond de’,  
*fànfɛ* ‘vers’,  
*jùkóro* ‘sous’,  
*kàlamà* ‘au courant de’,  
*kámà* ‘pour’,  
*kánna* ‘sur’,  
*kèrɛfɛ* ‘à côté de’,  
*kófɛ* ‘derrière’,  
*kósòn* ‘à cause de’,  
*kókàn* ‘à l’extérieur de’,

*kùndá* ‘du côté de, au niveau de ; selon’  
*kùnná* ‘sur’,  
*nɔ̃fɛ* ‘derrière’,  
*nɔ̃ná* ‘à la place de’,  
*ɲɛ́fɛ* ‘devant’,  
*ɲéna* ‘devant’,  
*ɲémà* ‘devant’,  
*ɲékɔ̃rɔ* ‘à la veille de’,  
*sánfɛ* ‘par-dessus’,  
*sɛ́nfɛ* ‘au cours de’,  
*sɛ́nkɔ̃rɔ* ‘parmi, au milieu de’.

Deux postpositions, *kósɔ̃n* et *kámà*, se rapprochent des postpositions composées par l’absence de compacité tonale, mais elles n’ont pas d’étymologies transparentes au niveau du bambara, ce qui les rend semblables aux postpositions simples.

L’écriture collée des postpositions composées a été recommandée dans les deux guides d’orthographe (Anonyme 1979: 20; Anonyme 1993: 17) ; à leur sujet voir également (Dumestre 2001). Cependant, dans les publications en bamanankan, on trouve souvent leur écriture séparée.

***Après une discussion animée, l’Atelier a confirmé l’écriture collée des postpositions composées.***

### 2.2.7. Les numéraux

Selon le Guide de transcription (Anonyme 1993), les numéraux, cardinaux comme ordinaires, s’écrivent séparément des noms précédents :

*mògɔ fila* ‘deux personnes’, *dón kò̀nɔ̃ntɔ̃nnan* ‘neuvième jour’, *Mùso jòli?* ‘Combien de femmes ?’ (*jòli* est un pronom-numéral).

Cependant, l’orthographe des numéraux de deuxième ordre formés avec *bí* n’avait pas été établie. Les dictionnaires publiés varient à ce propos ; ainsi, l’écriture collée est assumée dans (Dumestre 2011), et l’écriture séparée, dans (Anonyme 1980; Dukure & Baalo 2021). Dans (Bailleul 2007), on trouve les deux écritures en compétition.

Dans les textes publiés, on trouve les deux écritures, mais celle séparée est de loin plus fréquente. Un autre argument pour l’écriture séparée est l’absence de compacité tonale : *bí*, l’élément désignant la dizaine, porte un ton haut suivi d’un ton flottant bas, et la deuxième partie du numéral composé garde son ton lexical :<sup>8</sup>

---

<sup>8</sup> Un autre argument important a été évoqué par Klaudia Dombrowsky (communication personnelle) : il est possible de dire en bambara *bí cáman* ‘des nombreuses dizaines’, et les

*bî dúuru* 50, *bî kàntàntà* 90

*Après une discussion, l'écriture séparée pour ces numéraux a été retenue.*

### 2.2.8. Les formes redoublées

Dans le guide de la nouvelle orthographe (Anonyme 1993: 14), il a été recommandé de séparer les parties d'une forme redoublée avec un trait d'union. Cependant, on trouve en bamanankan des mots redoublés de deux types :

– ceux pour lesquels des mots non-redoublés dont ils seraient dérivés n'existent pas, ou les mots redoublés non-motivés. Le plus souvent, ils sont tonalement compacts. Ils peuvent appartenir aux différentes parties de discours. Il a été proposé dans (Konta & Vydrin 2014) d'écrire les mots redoublés non-motivés collés :

*kúlukulu* 'poulailler', *cógocogo* 'remuer dans l'eau', *sògɔsɔgɔ* 'tousser, la toux', *kàlakala* 'brindilles', *jàmajama* 'bricoles'.

– ceux pour lesquelles on trouve des formes non-redoublées dont ils ont été dérivés, comme *bòli* 'courir' pour *bòli-bòli* 'courir dans tous les sens', *tà* 'prendre' pour *tà-tà* 'prendre à plusieurs reprises', *pán* 'sauter' pour *pán-pán* 'sautiller'. Ils sont, en règle, tonalement non-compacts. Ces mots sont des « formes redoublées motivées » ; il a été proposé de les écrire avec des traits d'union. Les redoublés motivés sont surtout les verbes (à valeur de pluriaction, voir les exemples ci-dessus) et des numéraux à valeur distributive (on les écrit toujours en lettres) *kélen-kélen* 'par un', *fila-fila* 'par deux', *wóɔɔ-wóɔɔ* 'par six'.

L'écriture avec un trait d'union est observée même là où le verbe redoublé est doté d'un suffixe flexionnel :

Ù *bòli-bòlila* kà ò sògo tóbilenw dí m̀̀gɔ bée mà. 'Ils ont couru par-ci par-là et ont donné de cette viande cuite à tout le monde'.<sup>9</sup>

Ù ní jógɔncesiraw bée *tìge-tìgera* péwu. 'Les routes entre eux ont été toutes coupées (une par une)'.

On trouve quelques paires minimales des formes redoublées motivées et non-motivées, comme *sògɔ-sògɔ* 'percer par ci par là, trouer par ci par là' (de *sògɔ* 'percer') vs. *sògɔsɔgɔ* 'tousser, la toux'. Dans ces cas, l'orthographe différente (écriture collée vs. séparée par un trait d'union) permet de distinguer les sens.

---

adjectifs, d'après la règle générale, s'écrivent séparée des noms précédents. Je peux rajouter, dans la même logique, la possibilité de combiner *bî* avec le déterminant *yírìka* 'plusieurs'.

<sup>9</sup> Ici et dans ce qui suit, je donne des exemples phrastiques sans gloses. D'un part, les règles standards du glosage sont en contradiction avec la présentation des règles orthographiques (cela concerne surtout l'utilisation des traits d'union). D'autre part, l'audience de cet article est composée surtout de ceux qui connaissent bien la langue bambara et n'ont pas besoin des gloses.

***La proposition de l'écriture collée pour les formes redoublées non-motivées et avec un trait d'union pour les formes redoublées motivées a été retenue après une longue discussion.***

### 2.2.9. Noms composés tonalement non-compacts

En bamanankan, il y a des noms composés de types différents dont les parties constitutives ne sont pas tonalement compactes. Après une discussion très animée, il a été décidé de les écrire avec des traits d'union (8 voix contre 3 ; l'option alternative avait été d'écrire leurs composantes comme des mots séparés).

Parmi ces noms composés non-compacts, la majorité sont des conglomérés, c.-à.-d. des noms composés basés sur des constructions autres que celles des groupes nominaux (Dumestre 1983). La plupart des modèles des noms conglomérés ne sont pas tonalement compacts (les composantes gardant leurs tons d'origine), ces noms sont écrits avec des traits d'unions : *sìgi-ń-bólo-mà* 'personne tranquille ; tranquillité', *sé-ń-yèrê-kóro* 'autosuffisance', *ála-ká-mùrujan* 'arc-en-ciel', *ntòri-ká-sìgilan* 'espèce de champignon', *dùgutigi-y'-án-wéle* 'chemise sans manches'.

Cependant, dans un certain nombre des conglomérés, les tons originaux ne sont pas maintenus, et le mot appartient à un de deux types tonaux standards. Dans des cas pareils, on l'écrit collé : *sìginfè* 'le fait de vivre ensemble'.

En particulier, les conglomérés du type « V ka V » (verbe – marque de l'infinitif – verbe) ne sont pas tonalement compacts, on les écrit collés : *bàlakawuli* 'ganga de Gambie (oiseau)', *sòlikawuli* 'le fait de se lever de bonne heure'.

Il se trouve des cas où la composition nominale suit un modèle syntaxique basique (le plus souvent, une construction génitive, Nom + Nom), mais, contrairement aux attentes, le mot composé n'est pas tonalement compact. Dans de pareils cas, l'écriture avec un trait d'union est appliquée :

*tù-shènin* 'poule de rochers', *bá-wòlo* 'oie d'Egypte', *dàba-kàla* 'arbre à bois dur', *jí-sò* 'un animal aquatique'.

### 2.2.10. L'écriture des adverbes pré-verbaux

En bamanankan il y a une classe des adverbes qui précèdent les verbes et, le plus souvent, forment avec eux des unités tonalement compactes : *sèbekoro* 'sérieusement', *dàmateme* 'excessivement, etc. ; voir (Dumestre 1987: 484–490; Dumestre 2003: 289–292).

A part cela, il existe un mécanisme productif de conversion régulière des noms en adverbes préverbaux à valeur comparative, comme *fàli* 'âne' dans *Û y'à fàlibugò* 'Ils l'ont frappé comme un âne' (Vydrin 2012; Vydrin 2019: 379–382).

Dans le Guide de transcription (Anonyme 1993: 13), une écriture séparée a été recommandée pour les adverbes préverbaux (y compris ceux dérivés des noms).

*Après une longue discussion, l'Atelier a recommandé une écriture collée pour tous les types des adverbes préverbaux (y compris ceux provenant des noms) : Án b'ò sèbekɔrɔɔn 'Nous le savons parfaitement', Û ye dùgu ìn dàgakolonci 'Ils ont écrasé ce village comme une vieillealebasse'.*

### 2.2.11. Écriture collée vs séparée pour certains mots/expressions fréquentes

2.2.11.1. On trouve en bamanankan des séquences des phonèmes identiques qui expriment des sens différents. Parfois ces séquences portent des tons différents, et parfois elles ne se distinguent même pas de ton. L'écriture séparée ou collée permet de les distinguer (cependant, dans la pratique écrite du bamanankan, cette différence n'est pas toujours respectée). Énumérons certains de ces cas.

– *mùnná* 'pourquoi ?' vs. *mùn ná* 'dans quoi?', ex. :

*Í ye à kàlan mùnná ?* 'Pourquoi l'as-tu enseigné ?' vs.

*Í ye à kàlan mùn ná ?* 'Qu'est-ce que tu lui as enseigné ?'

– *kójugu* 'très' vs. *kó jugu* 'affaire grave, crime, mauvaise pensée', ex.:

*À ka dàgaw ka nì kójugu!* 'Ses canaris sont très bons'.

*Ò tuma` lá, kó jugu` donna ù dùsukun` ná.* 'A ce moment, une mauvaise pensée est entrée dans leurs cœurs'.

– *k'à sòrɔ, kàsɔrɔ* (l'Atelier a décidé de retenir les deux écritures) 'bien que, malgré' (valeur de concession) vs. *ka sòrɔ* 'puis' (valeur consécutive), ex. :

*Ò wáatiw lá, mùso` tòn bé sàñ bì náani ké dénwoło` lá kàsɔrɔ/k'à sòrɔ à má bàna mìnɛ.* 'A cette époque-là, une femme accouchait pendant quarante ans, sans tomber malade'.

*Í be finfin` ké ò kónɔ, kà sòrɔ kà cèncèn` dá ò kàn...* 'Mets du charbon là-dedans, puis rajoute du sable...'

– *kàbán* 'déjà' vs. *kà bán* ou *k'à bán* 'complètement', ex. :

*À binaake nìn, à yèrè sàra kàbán.* 'Son oncle maternel lui-même est déjà mort'.

*Ò y'à sòrɔ mùso` tílala à ká kó` lá kà bán péwu.* 'Il s'est avéré que la femme a fini son affaire complètement'.

– *jémà* 'devant', une postposition composée vs. *jé` mà*, dans l'expression *à jé` mà* 'comme il faut' :<sup>10</sup>

*À b'í n'à fɔ fèn` dó tòn bé kálo` jémà.* 'C'était comme s'il y avait eu quelque chose devant la lune'.

*N' b'ò dón à jé` mà.* 'Je le connais très bien, comme il faut'.

<sup>10</sup> Dans l'expression *à jé` mà* 'comme il faut', la particule de focalisation *dè* peut être insérée entre les mots *jé* et *mà*, ce qui justifie leur écriture séparée (malheureusement, dans les publications en bamanankan, cette distinction est peu respectée, et le plus souvent, on écrit *jé` mà* en un seul mot, \**jema*). Il faut mentionner également l'adjectif *jéma* 'pourvu des yeux' qui se diffère de la postposition *jémà* par son ton.

## 2.2.11.2. Quelques cas individuels problématiques

L'Atelier a fait la discussion sur quelques mots et expressions pour lesquelles on observe des divergences dans la pratique écrite du bamanankan, cf. Tableaux 2 et 3.

Sens	Forme recommandée	Forme déconseillée
'quelques'	<i>dámado</i>	* <i>dáma dó</i>
'un autre'	<i>dówere</i>	* <i>dó wére</i>
'depuis'	<i>kàbíni</i>	* <i>kàbí ní</i>
'très, beaucoup'	<i>kósebe, kósèbe</i>	* <i>kó sèbe</i>
'bien, comme il faut'	<i>kópùman</i>	* <i>kó pùman</i>
'de nouveau'	<i>kókura</i>	* <i>kó kúra</i>
'quand, lorsque'	<i>mînké, mînkéni</i>	* <i>mîn ké, *mîn kéni</i>
'là-bas'	<i>yènnínnò, yènnò</i>	* <i>yen nin nò</i> ou * <i>yènní nò, *yen nò</i>

Tableau 2. Formes liées recommandées

Sens	Recommandé	Déconseillé
'à l'oubli de'	<i>à jìne` mà</i>	* <i>a jìnema</i>
'derrière le fleuve, sur l'autre rive de fleuve'	<i>bá` kó</i>	* <i>bakò</i>
'sous la main'	<i>bólo kórò</i>	* <i>bolokòrò</i>
'faire une intrusion; mettre fin, annuler'	<i>dàn kárí<sup>11</sup></i>	* <i>dankari</i>
'par terre'	<i>dùgu` mà</i>	* <i>duguma</i>
'jusqu'ici'	<i>háli bì</i>	* <i>halibi</i>
'pardonner'	<i>hàke tó</i>	* <i>haketo</i>
'comme'	<i>í n`à fò</i>	* <i>inafò, *i na a fò</i>
'comme, comme si'	<i>í kó</i>	* <i>iko</i>
's'ingérer'	<i>k`í sèn` dòn</i>	* <i>k`i sendon</i>
'se suicider'	<i>k`í yèrê fâga</i>	* <i>k`i yerefaga</i>
'au hasard'	<i>kùn fê</i>	* <i>kunfê</i>
'grâce à'	<i>nùn mà</i>	* <i>nunma, *numa</i>
'au cours de'	<i>sèn fê</i>	* <i>senfê</i>
'en marche, en cours, sur pied'	<i>sèn kàn</i>	* <i>senkan</i>

Tableau 3. Formes séparées recommandées

<sup>11</sup> Très bizarrement, une habitude s'est établie dans les publications en bamanankan d'écrire collée l'expression *dàn kári* 'faire une intrusion ; mettre fin, annuler'. Cependant, l'emploi de cette expression dans une construction transitive en aspect accompli (perfectif) exige la marque prédicative *ye* qui signale clairement qu'il s'agit d'une construction verbale transitive. Ainsi, on trouve dans Kibaru No 356 : *Banbaganciw ye dankari Ameriki jamana la* 'Les terroristes ont attaqué les États-Unis'. La seule interprétation possible est que *dan* soit un complément d'objet directe (*dàn* 'limite'), et *kári* soit le verbe ; l'écriture collée s'avère donc erronée.

Il y a une expression, [kàà mäsòrò] ‘parce que’, pour laquelle les experts ont retenu les deux écritures : *k’a mäsòrò* (écriture séparée) ou *kàmasòrò* (écriture liée).

### 2.2.12. Noms propres

Les noms propres sont écrits avec la première lettre en majuscule. Les groupes suivants sont considérés comme des noms propres (et s’écrivent donc avec les majuscules ; cette liste n’est pas exhaustive) :

- prénoms de gens (et noms des animaux): *Áminàta, Sékù*,
- nom de familles, désignations de membres de familles/clans : *Jàra, Kùlubàli, Kàmisoko, Kùlibalilaka* ‘membre de la famille / du clan Coulibaly’,
- les habitats des familles : *Jàrala, Kùlubalila*,
- noms des localités (toponymes) : *Bàmakò, Líberiya*,
- équipes, noms de compagnies : *Árajomali* ‘Radio Mali’, *Sótarama* ‘SOTRAMA’,
- noms de monuments : *Kònòba, Bugiba, Màliba, Sìgi, Sàmaba*.

Si un nom propre consiste en plusieurs mots, chaque mot est écrit avec une lettre majuscule, ex. :

- Promenade des Angevins : *Anzekaw Tògòlasiraba*,
- Jardin du maire de Bamako : *Bamakò Meri Tògòlafijñemineyòrò*,
- Hôpital Gabriel Touré : *Gabiriyeli Ture Dògòtòròsoba*,
- Square Patrice Lumumba : *Patirisi Lumunba Tògòlakene*,
- Rond-point Kwamé Krouma : *Kuwame Kuruma Tògòlakurubò*.

Ne sont pas vus comme des noms propres les catégories suivantes de noms (qui sont donc écrits avec des lettres minuscules) :

- jours de la semaine, les mois: *àraba* ‘mercredi’, *zánwuye* ‘janvier’,
- noms de groupes ethniques, des nations, des habitants de lieux : *sínefò* ‘Sénoufo’, *séguka* ‘habitant/originaire de Ségou’,
- noms des langues : *sínefòkan*.

### 2.2.13. Mots étrangers

Pour les mots étrangers non-adaptés (surtout des noms propres), l’Atelier a décidé de continuer la pratique actuelle de l’adaptation de l’orthographe de ces mots au modèle bamanankan : *Berinari* ‘Bernard’, *Zozefini* ‘Josephine’, *Yesu Kirisa* ‘Jesus Christ’, *Abuduli Kaderi* ‘Abdoul Kader’; *Zerari Dumési* ‘Gérard Dumestre’, *Betowen* ‘Beethoven’, *Digoli* ‘de Gaulle’, *Olandi* ‘Hollande’, *Merikeli* ‘Merkel’.

### 2.2.14. Les sigles et acronymes

**Les sigles et acronymes** provenant des langues autres que bamanankan s’écrivent dans le respect de la prononciation du bamanankan sans placer ni des

points ni des tirets. On accompagne ces formes par le sigle originel (français, anglais, etc.) : *Inesiko* (UNESCO), *Inisɛfu* (UNICEF).

### **2.3. Propositions de l'article (Konta & Vydrin 2014) non retenues par l'Atelier**

Certaines propositions formulées dans l'article de Konta & Vydrin (2014) ont été rejetées par les participants de l'Atelier, en particulier :

– La réintroduction de *gw* (dont l'utilisation a été déconseillée dans les guides de transcription). En fait, *gw* représente un phonème à part du bamanankan, différent de /g/, le problème est en son instabilité élevée dans les dialectes et même en bambara standard. L'Atelier a décidé de maintenir la recommandation négative par rapport à l'utilisation de *gw*.

– Il a été proposé de ne pas utiliser l'élision vocalique finale pour les verbes (c.-à-d., ne pas écrire *À d'à mà* 'Donne-le lui', *Án bór'à lá* 'Nous sommes sortis de cela', et utiliser les formes complètes : *À dí à mà*, *Án bóra à lá*). L'Atelier a décidé de ne pas introduire cette restriction, pourvu que cela ne perturbe la compréhension.

### **3. Sélection des variantes standards des lexèmes**

Les débats sur les règles ont pris deux jours de travail de l'Atelier, ce qui a été suivi de presque trois semaines de travail sur la liste lexicale. La procédure était la suivante : la liste des variantes à discuter, avec toutes les données statistiques, était projetée sur l'écran. Chacun était invité à évaluer les variantes et à suggérer, laquelle d'entre elles devrait être prise pour standard. Si les avis divergeaient, on passait au vote aux mains levées. Parfois, à la demande de participants, certains lexèmes étaient rediscutés et revotés.

Assez souvent on tombait sur des lexèmes que certains experts ignoraient ; dans ces cas, seuls les participant familiers avec le lexème contribuaient à la discussion et au vote.

Les résultats de ce travail se présentent dans la nouvelle version du dictionnaire Bamadaba en ligne (Bailleul et al. 2011), et les nouvelles versions du Corpus Bambara de Référence seront annotées en tenant compte des choix orthographiques dans le dictionnaire. Il serait souhaitable qu'un dictionnaire orthographique bambara à l'usage des enseignants et élèves maliens soit publié au Mali en version papier, mais cela dépend des instances chargées de la politique linguistique du pays, et cette tâche semble difficilement réalisable dans la situation actuelle au pays.

Les tableaux utilisés pendant le travail de l'Atelier seront mis en ligne sur le site du Corpus Bambara de Référence et du dictionnaire Bamadaba ; dans ces tableaux on trouvera les données statistiques pour chaque variante de chaque lexème, et les



indicateurs désignant les décisions prises (*r* – variante recommandée, *t* – variante tolérée, *e* – variante à éliminer).

Il ne s’agit pas de discuter ici tous les lexèmes traités, mais je présenterai quand même quelques cas typiques auxquels les participants de l’Atelier se sont confrontés.

### 3.1. Les variantes « tolérées »

Comme il a été déjà dit, le but d’un dictionnaire orthographique est la normalisation (autrement dit, la standardisation) de la langue écrite bambara. La normalisation sous-entend l’élimination de la variabilité, ou, au moins, sa minimalisation. Cependant, le nombre des variantes « tolérées » s’avère assez élevé, car on a fait recours à ce statut souvent en quête de compromis. Il y a eu même des cas où, pour des raisons multiples, on a attribué ce statut à plus d’une variante, ex. :

*hére r, héra t, héere t* ‘paix, bonheur’,

*lìngilanga r, lèngelanga t, jìngijanga t* ‘se dandiner’,

*kàmali r, kàmani t, kàmari t* ‘rassembler’.

Le statut « tolérée » a été souvent attribué à toutes les variantes des adverbes expressifs (ou autres mots d’origine idéophonique), car ces variantes peuvent exprimer des nuances sémantiques, parfois évasives. Ex. :

*kélenwu r, kéwu t, kélewu t, kéwu t, káwu t* ‘bruit de casse’

### 3.2. Des variantes recommandées malgré les données statistiques

L’idée originale avait été de choisir les variantes standards en s’appuyant surtout sur les données statistiques, c.-à-d. sur leurs fréquences dans le Corpus Bambara de Référence. Cependant, lors des discussions au sein de l’Atelier, d’autres arguments ont été considérés comme plus consistants. Quelques exemples sont présentés dans le Tableau 4.

Trad.	Variante	Statut	Corpus non-désamb.	Corpus désamb.	Auteurs	Documents
écraser	<i>cónkən</i>	e	137	127	39	5
	<i>cóngɔ</i>	r	30	29	24	11
	<i>cónkɔ</i>	t	20	20	12	4
créature	<i>dánfən</i>	t	351	350	178	44
	<i>dáfən</i>	r	252	249	119	35
hameçon	<i>dóolen</i>	r	3	3	2	2
	<i>ndóolen</i>	e				
	<i>dólen</i>	t	56	56	17	9
	<i>ndólen</i>	e	1	1	1	

Trad.	Variante	Statut	Corpus non-désamb.	Corpus désamb.	Auteurs	Documents
	<i>dólen</i>	e	15			
courber	<i>dù</i>	e	5087	65	52	25
	<i>dùn</i>	r	13832	4	4	4
veuvage	<i>firiya</i>	t	64			
	<i>filiya</i>	r	52	21	3	3
	<i>fùruja</i>	e	3			
	<i>fiiya</i>	e	2			

Tableau 4. Variantes recommandées choisies malgré les données statistiques

Les raisons de ces choix pouvaient être différents. Ainsi, la variante *dáfen* ‘créature’ a été retenue pour donner suite au choix de la variante *dá* pour le verbe ‘créer, fabriquer’ (par défaut, la même forme doit être choisie dans les mots composés et dérivés, sauf s’il y a de très forts arguments pour le contraire). La variante *dóolen* pour ‘hameçon’ a été retenue à cause de sa voyelle longue, un trait distinctif ancien. Et dans des nombreux cas, l’argument principal a été l’intuition linguistique de la majorité des participants.

### 3.3. Les variantes recommandées ou tolérées absentes de la liste originale

Il y a eu des cas où les variantes phonétiques de la liste originale (provenant du dictionnaire Bamadaba) ont été complétées, lors de la discussion, par d’autres variantes, et la prépondérance (le statut « tolérée » ou même « recommandée ») a été donnée aux variantes rajoutées.

Il s’agit, en fait, des lacunes dans la version du dictionnaire Bamadaba utilisée comme base pour la liste des variantes lexicales, lacunes comblées au cours de la discussion. D’ailleurs, la plupart des variantes (mais pas toutes) rajoutées à la liste apparaissent dans le Corpus Bambara de Référence. Voici quelques cas.

*dèndegelu* r ‘épier, regarder en tendant les yeux et en se cachant’, les autres formes, *dègedegelu*, *dègedegeli*, *dèndegeru*, *dèndege*, ont été « rejetées ». C’est un lexème peu fréquent dans les textes (seules 2 occurrences dans le Corpus, dans les deux cas, c’est la variante *dègedegeli*).

*dèmedeme* r ‘se démener ; aider à marcher’, une variante non-représentée dans le Corpus où seule la forme *dèmedeme* apparaît. Cependant, les experts ont voté pour la recommandation de la variante avec les voyelles longues.

*álajì* r, *lájì* t ‘hadji’. Dans la liste originale (et dans la version de Bamadaba, ce lexème avait trois variantes : *élihajì*, *éli hajì*, *hájì* ; les experts en ont rajouté trois autres. Dans le Tableau 5, les statistiques des occurrences des variantes dans le

Corpus sont présentées. Comme on voit, les deux variantes les plus fréquentes ont été rejetées par les experts, et la variante « tolérée » n'est pas du tout attestée dans le Corpus.

Variante	Statut	Corpus non-désamb.	Corpus désamb.	Auteurs	Documents
<i>elihaji</i>	e	60	59	43	24
<i>haji</i>	e	145	1	1	
<i>eli haji</i>	e	103	103	13	12
<i>álajì</i>	r	49			
<i>lájì</i>	t	0			
<i>éłajì</i>	e	0			

Tableau 5. Statistiques des occurrences pour *álajì* ‘hadji’

*fisika* r, *fisinka* e, *finsan* e ‘cuire au bain-marie’ est un mot peu fréquent ; on a retenu la variante qui a été familière au moins à quelques participants de l'Atelier.

*júra* r ‘visiter un lieu saint’ est un autre mot peu fréquent. Les deux variantes de la liste originelle, *júura* et *jóra*, ont été rejetés par l'Atelier, au profit de la forme à voyelle brève.

*kéreke* ‘selle’ (voir Tableau 6) est un ancien emprunt, probablement au soninké. Aux quatre variantes de la liste originale (*kírike*, *kírike*, *kílike*, *kérike*) les experts ont rajouté deux, dont une, *kéreke*, a été retenue comme standard (contrairement aux statistiques des occurrences dans le Corpus), tandis qu'une des formes originales, *kírike*, a obtenu le statut de « tolérée ».

Variante	Statut	Corpus non-désamb.	Corpus désamb.	Auteurs	Documents
<i>kírike</i>	t	60	59		18
<i>kírike</i>	e	73	72		17
<i>kílike</i>	e	6	6		1
<i>kérike</i>	e	7	6		5
<i>kéreke</i>	e				
<i>kéreke</i>	r	12			

Tableau 6. Statistiques des occurrences pour *kéreke* ‘selle’

*éreni* ‘hernie’ est un emprunt français. Dans la liste originale, on trouve trois variantes : *éreni*, *érini*, *éreni*, dont une seule, *éreni*, apparaît dans le Corpus. A cette

occasion, on peut mentionner le principe que les experts ont décidé d'appliquer aux emprunts français récents est d'éviter la variabilité (pas de variantes « tolérées » !).

On trouve également des cas où une variante rajoutée à la liste par les experts est, suite à la discussion, marquée comme « tolérée ». Il s'agit aussi de lacunes importantes remplies au cours du travail de l'Atelier. Ex. :

*dádè* r, *dájè* t, *dádèn* e 'se taire'. La variante *dájè* avait été absente, et on ne la trouve pas dans le Corpus. D'ailleurs, la variante rejetée *dádèn* est très bien représentée (68 occurrences dans le Corpus, dont 36 dans le sous-corpus désambiguïsé ; 16 textes, 6 auteurs).

*fěngɛɲɛ* r, *pěngɛɲɛ* t, *fěngɛngɛ* e 'derrière, fesses'. La variante *pěngɛɲɛ* a été rajoutée par les experts au cours de discussion.

### 3.4. Lexèmes que les experts ignorent

Le bambara étant une grande langue, avec beaucoup de dialectes et variantes régionales, il n'est pas étonnant qu'un nombre sensible de lexèmes figurant dans les grands dictionnaires s'est avéré inconnu de nos experts. Sans doute, ce sont des mots régionaux ou obsolètes ; il s'agit souvent des noms des espèces biologiques que les citadins ignorent très naturellement. Voici quelques exemples :

*dàga-sìgi-dàga-lá* 'bouton blanc' (Leucas martinicensis), *dònká* 'petit tambour de griot', *gága* 'panier de cola', *làfa* 'bonnet', *nòɔrɛ* 'bourses', *ntěnkisɛdabilen* 'arbre érythrine (Erythrina senegalensis)', *sòmi* 'acte sexuel', *wòyo*, *wòyo* 'faire fondre', *ngíran* 'oiseau échasse blanche'.

Quand les lexèmes de ce type ont des variantes phonétiques, le seul moyen de classifier leurs variantes reste les données statistiques ; et si ces données manquent (le mot n'apparaît pas dans le Corpus, ou il n'y apparaît que très rarement), toutes les variantes ont reçu le statut soit « recommandé », soit « toléré ».

D'autres mots ont été connus d'un seul expert, voir de deux. Ainsi, tout le monde a ignoré les mots suivants, sauf Dokala Diarra (originaire du Bélé Dougou ; la liste n'est pas exhaustive) : *dúngùru* 'planchette sur le linteau de la porte', *hávule* 'gourmandise', *nòmi* 'bancaal', *sěli* 'bouder', *tìntɔ*, *tìntɔn* 'jeune initié'.

D'autres mots n'ont été connus que d'Issiaka Ballo, originaire du Baninko, par exemple: *fúse* 'gris', *kòlonkòlon* 'tambour rituel', *kònkònnìn* 'silure', *tórilan* 'marteau'.<sup>12</sup>

L'information « négative » sur l'usage de mots (c.-à-d., le fait qu'ils soient peu connus et peu utilisés dans la communauté linguistique) est rarement présente dans les dictionnaires. Et pourtant, cette information est d'une grande valeur pour la compréhension de la dynamique de la langue.

---

<sup>12</sup> Curieusement, *fúse* et *kònkònnìn* sont absents du dictionnaire (Dukure & Baalo 2021).

### 3.5. D'autres cas compliqués

Lors des discussions de l'Atelier, beaucoup d'autres sujets lexicographiques ont été abordés, et la présence d'un groupe représentatif d'experts-locuteurs natifs a permis de trouver des réponses. Par exemple, on a établi ou confirmé que certains mots ont des variations tonales (régionales ou autres ; la distribution des variantes tonales demande une étude beaucoup plus approfondie). Voici quelques exemples :

*náganagalèn* (Ségou), *náganagalen* (Bamako) 'hirondelle' ;

*mà* ou *má* 'lamantin', ton bas pour tous les experts, sauf Ibrahima Diawara ;

*mà* ou *maa* (atonal) marque de l'optatif : pour la majorité d'experts : une voyelle brève, ton bas ; pour une minorité : une voyelle longue, ton variable ;

*sàgon* (Bélédougou), *sáfo* (tous les autres) 'fade' ;

*kókɔri* (Issiaka Ballo), *kòkɔri* (tous les autres experts) 'instrument recourbé pour évider les Calebasses'.

Il est vrai que l'information sur la tonalisation des mots n'a pas d'incidence directe sur l'orthographe bambara (car les tons n'y sont pas notés), mais elle est très importante pour la lexicographie et les études tonologiques.

D'autres cas compliqués ont trouvé des réponses également. Ainsi, il existe en bambara deux formes pour 'cochon', *lè* et *jè*, les deux ayant, sans doute, la même origine.<sup>13</sup> Pour une moitié des participants de l'Atelier, *lè* désigne le cochon domestique, et *jè*, le cochon sauvage ; on peut donc les considérer comme des lexèmes différents (apparemment, cette distinction est plus typique de la partie est de l'aire linguistique bambara). Pour les autres, *lè* et *jè* ont le même sens, il s'agit donc des variantes du même lexème.

## 4. Conclusions

Cet article n'a aucune ambition théorique ; son but est juste de présenter l'état actuel du projet d'un dictionnaire orthographique bambara, et les circonstances dans lesquelles le travail sur ce dictionnaire se déroule.

Il faut mentionner également un aspect important, souvent évoqué en cette liaison : ce dictionnaire orthographique bambara, comment s'inscrit-il dans le processus d'harmonisation des orthographe des variétés mandingues différentes ?<sup>14</sup>

L'idée d'une telle harmonisation est débattue depuis les années 1960. En fait, il s'agit surtout de l'harmonisation des variétés appartenant au sous-groupe mandingue de l'Est : le bambara du Mali, le maninka de Guinée, le dioula de Côte d'Ivoire et le

---

<sup>13</sup> Selon toute apparence, *lè* est une forme d'origine maninka, et *jè* est la vraie forme bambara ; les deux proviennent de la forme Proto-Mandingue que je reconstruis comme \**dê*.

<sup>14</sup> Je ne touche pas ici aux relations entre l'écriture latine et le N'ko, c'est un problème à part. Seule l'écriture latine est concerné par cet article.

dioula du Burkina Faso (il s'agit, bien évidemment, des variantes « standards » de chaque pays, les dialectes locaux restant quasiment sans écriture). Les variétés du sous-groupe mandingue de l'Ouest sont linguistiquement trop distants (il suffit de mentionner que leur système vocalique compte 5 phonèmes, contre 7 voyelles en mandingue de l'Est), et en fait, on n'a jamais entrepris d'efforts sérieux pour rapprocher l'orthographe du mandinka du Sénégal et de Gambie avec les variétés orientales du mandingue.

L'harmonisation des variétés orientales du mandingue a apporté certains résultats. Leurs alphabets actuels sont pratiquement identiques ; même s'il y a certaines conventions divergentes (comme l'utilisation de la double voyelle dans l'orthographe dioula du Burkina Faso pour signaler l'article tonal), elles n'empêchent pas (ou très peu) la lisibilité mutuelle des textes.

Cependant, une standardisation entre le bambara, le dioula de RCI, le dioula du BF et le maninka de la Guinée aux niveaux lexical et grammatical ne me semble pas réalisable. Le travail fait par l'Atelier en mai-juin 2022 a montré les difficultés d'une standardisation même au niveau du bambara seul ; et si on y rajoute les trois autres grandes variétés du mandingue oriental, cela risque de rendre la tâche impossible, et ses résultats peu convaincants.

### Références

- Anonyme. 1979. *Guide de transcription et de lecture du bambara*. Bamako: DNAFLA.
- Anonyme. 1980. *Lexique bambara-français (bamanan kan ni tubabu kan danyègafè)*. Bamako: Imprimerie DNAFLA.
- Anonyme. 1993. *Guide de transcription et de lecture de la langue bambara*. 2nd edn. Bamako: Imprimerie DNAFLA.
- Bailleul, Charles. 2007. *Dictionnaire Bambara-Français*. 3rd edn. Bamako: Donniya.
- Bailleul, Charles, Artem Davydov, Anna Erman, Kirill Maslinsky, Jean-Jacques Méric & Valentin Vydrin. 2011. *Bamadaba : Dictionnaire électronique bambara-français, avec un index français-bambara*. <http://cormand.huma-num.fr/bamadaba.html>.
- Dukure, Mamadu F. & Isiyaka Baalo. 2021. *Bamanankan danegafè*. Bamako: EDIS.
- Dumestre, Gérard. 1983. Les conglomérés en bambara. *Mandenkan* 6. 37–53.
- Dumestre, Gérard. 1987. *Le bambara du Mali : Essai de description linguistique*. INALCO. Paris.
- Dumestre, Gérard. 2001. Les locutions postpositives du bambara. *Mandenkan* 37. 1–8.
- Dumestre, Gérard. 2003. *Grammaire fondamentale du bambara*. Paris: Karthala.
- Dumestre, Gérard. 2011. *Dictionnaire bambara-français suivi d'un index abrégé français-bambara*. Karthala. Paris.
- Kilgarriff, Adam. 1997. Putting frequencies in the dictionary. *International Journal of Lexicography* 10(2). 135–155.

- Konta, Mahamadou & Valentin Vydrin. 2014. Propositions pour l'orthographe du bamanankan. *Mandenkan* 52. 22–54. <https://doi.org/10.4000/mandenkan.318>.
- Vydrin, Valentin. 2012. Predglagol'nyje narechija v bamana (Предглагольные наречия в бамана) [Pre-verbal adverbs in Bamana]. In Valentin Vydrin & Alexander Zheltov (eds.), *Mezhdú Nigerom i Kongo: zametki na pol'akh. K 60-letiju Konstantina Pozdniakova* (Между Нигером и Конго: Заметки на полях. К 60-летию Константина Позднякова) [*Between the Niger and the Congo: Field notes. To the 60th anniversary of Konstantin Pozdnyakov*], 79–93. St. Petersburg: Nestor-Istoria.
- Vydrin, Valentin. 2019. *Cours de grammaire bambara*. Paris: Presses de l'INALCO.
- Vydrin, Valentin. 2020a. Featural foot in Bambara. *Journal of African Languages and Linguistics* 41(2). 265–300. <https://doi.org/10.1515/jall-2020-2012>.
- Vydrin, Valentin. 2020b. Vowel elision and reduction in Bambara. *Italian Journal of Linguistics* 32(1). 103–124. <https://doi.org/10.26346/1120-2726-150>.
- Vydrin, Valentin, Kirill Maslinsky, Jean-Jacques Méric & Andrij Rovenchak. 2011. *Corpus Bambara de Référence*. <http://cormand.huma-num.fr/index.html>.
- Vydrin, Valentin & Jean-Jacques Méric. 2020. Corpus-driven Bambara spelling dictionary. In *Computational Linguistics and Intellectual Technologies Papers from the Annual International Conference "Dialogue" (2020)*, vol. 19, 1180–1187. Moscow: Russian State University for the Humanities. <https://doi.org/10.28995/2075-7182-2020-19-1180-1187>.

*Valentin Vydrin*

### **Vers un dictionnaire orthographique bambara**

L'article présente les résultats d'un projet de standardisation de l'orthographe bambara sur la base d'un corpus annoté des textes (Corpus Bambara de Référence). A la première étape, pour tous les lexèmes du dictionnaire Bamadaba ayant des variantes graphiques, les statistiques d'occurrences détaillées dans le Corpus ont été relevées. Puis, ces variantes ont été triées, ce qui a permis d'établir l'écriture normative pour la majorité des lexèmes variables. A l'étape finale, les lexèmes à formes variables pour lesquelles les données statistiques avaient été considérées insuffisantes ont été discutés par un groupe d'experts représentant des instances linguistiques maliennes différentes, et les variantes normatives ont été sélectionnées. En même temps, les règles d'orthographe bambara ont été revisitées et précisées là où cela s'est avéré nécessaire.

**Mots clé :** dictionnaire orthographique, langue bambara, linguistique de corpus, normalisation de langue

Valentin Vydrin

### **Toward a Bambara spelling dictionary**

This paper presents the results of a project of standardization of the Bambara orthography on the basis of an annotated corpus of texts (The Bambara Reference Corpus). First, all lexemes in the Bamadaba dictionary which had spelling variants were checked for their frequencies in the Corpus. Then the variants were sorted, and standard graphic variants for a great number of lexemes were established. As a final step, a list of lexemes for which statistics were insufficient was drawn up, and a group of experts representing different Malian linguistic institutions discussed this list and selected the variants which could be accepted as standard. At the same time, general spelling rules of Bambara were revisited and edited wherever necessary.

**Key words:** spelling dictionary, Bambara language, corpus linguistics, standardization of language

*Валентин Феодосьевич Выдрин*

### **Создание орфографического словаря бамана**

В статье излагаются результаты проекта по стандартизации орфографии языка бамана на основе корпуса аннотированных текстов (Справочного Корпуса Бамана). На первом этапе исследования все лексемы из словаря Бамадаба, имеющие графические варианты, были проверены в отношении частоты употребления каждого варианта в Корпусе. Затем была проведена сортировка этих вариантов на основании частот их употребления, и для большинства лексем была выбрана стандартная форма написания. Одновременно был составлен «остаточный список» лексем, для которых статистические данные оказались недостаточными для принятия решений о стандартном графическом варианте. Последним этапом стало рабочее совещание экспертов, представляющих различные малийские учреждения, в которых осуществляется изучение национальных языков. Эксперты обсудили лексемы из «остаточного списка» и выбрали стандартный вариант для каждой из них. Помимо этого, были заново рассмотрены правила орфографии бамана, и в некоторых случаях было предложено внести в них изменения.

**Ключевые слова:** орфографический словарь, язык бамана, корпусная лингвистика, стандартизация языка