

Vers une lexicographie mandingue sur la base de grands corpus annotés

Valentin Vydrin

INALCO — LLACAN — St. Petersburg State University

vydrine@gmail.com

1. Introduction

Depuis une dizaine d'années, des corpus annotés de textes écrits ont été créés pour plusieurs langues mandé, dont les plus volumineux sont ceux pour le bambara (plus de 11 millions de mots en avril 2020) et le maninka de Guinée (plus de 3,5 millions de mots), cf. plus en détail (Vydrin, Maslinsky & Méric 2011; Vydrin 2013; Vydrin 2014; Vydrin, Rovenchak & Maslinsky 2016).

L'annotation de ces corpus est effectuée avec les dictionnaires électroniques Bamadaba (Bailleul et al. 2011) pour le Corpus Bambara, et Malidaba (Vydrin 2018) pour le Corpus Maninka. Ces deux dictionnaires diffèrent beaucoup par leur degré d'élaboration. En avril 2020, le Bamadaba comptait plus de 13 150 entrées dans la partie principale (en outre, il comporte des annexes : dictionnaires des toponymes, des noms claniques et des prénoms, des termes encyclopédiques). Le Bamadaba est disponible sur le site du Corpus Bambara de Référence, il est constamment complété. Les sources de nouvelles données lexicographiques sont surtout les textes préparés pour le sous-corpus désambiguïsé et, d'autre part, d'autres dictionnaires, surtout, (Vydrine 1999; Dumestre 2011).

Le dictionnaire Malidaba (maninka–français–anglais–russe) a été mis en ligne¹ au printemps 2018. En avril 2020, il comptait un peu moins de 7000 entrées dans la partie principale ; de nouvelles données y sont intégrées constamment. Si l'on compare le Malidaba avec le Bamadaba, il est beaucoup moins homogène : certaines entrées sont assez élaborées, d'autres, surtout vers la fin du dictionnaire, le sont moins. Quoi qu'il en soit, c'est le premier dictionnaire bilingue (ou plutôt quadrilingue) maninka de taille moyenne avec notation tonale, l'information grammaticale et la polysémie plus ou moins élaborée, et avec la possibilité de navigation par les mots maninka et par les équivalents français, anglais et russes.

¹ Sur le site du Corpus Maninka de Référence, <http://cormand.huma-num.fr/cormani/>.

Si l'on compare le Bamadaba avec le dictionnaire de Dumestre (2011), on dirait qu'ils sont en complémentarité. Le dictionnaire de Dumestre dépasse le Bamadaba pour le nombre d'entrées (15 850) et souvent pour l'élaboration de la polysémie. Le Bamadaba gagne en structuration de la polysémie et en notation grammaticale, en ce qui concerne surtout les valences verbales. En plus, le Bamadaba a l'avantage d'être un dictionnaire électronique : il est régulièrement mis à jour, et, ce qui est très important, il a un accès direct au Corpus Bambara de Référence (chaque lexème comporte un lien à un concordancier généré à partir du sous-corpus désambiguïsé).

On peut dire que la lexicographie mandingue a déjà profité des projets des grands corpus annotés des langues mandingues. Cependant, il faut avouer qu'à ce stade le potentiel lexicographique de ces corpus n'est exploité qu'à un degré insignifiant. L'existence des corpus annotés comportant des millions de mots est un grand atout qu'il serait tentant d'utiliser pour créer des dictionnaires dérivés des corpus numériques (angl. « corpus-driven dictionaries »), en s'appuyant sur l'expérience de la lexicographie moderne. Un tel dictionnaire représenterait le vocabulaire d'une langue d'une façon stéréoscopique, ce qui permet d'observer chaque lexème de sorte qu'on voie plus clairement son fonctionnement dans la langue.

La littérature sur la lexicographie de corpus étant très riche,² je ne pense pas qu'il soit nécessaire de présenter ici les approches élaborées dans le cadre de cette discipline. Au lieu de cela, je chercherai à démontrer le potentiel du Corpus Bambara de Référence dans le domaine de la lexicographie avec l'exemple de deux lexèmes, le verbe *bánban* et le nom *jàhadi*.

2. Bánban

2.1. *Bánban* dans les dictionnaires « classiques »

Prenons pour le point de départ la présentation de ce lexème dans les grands dictionnaires bambara disponibles.

Chez Dumestre (2011)³ ce lexème apparaît en deux variantes phonétiques, *bánban* et *báman*, et ses sens sont présentés en liste par les équivalents français suivants :

² Il suffit de mentionner des ouvrages classiques (Kilgarriff 1997; Kilgarriff, Rundell & Dhonnchadha 2006; Hanks 2012).

³ Dans ce qui suit, codé comme DUM.

s'évertuer, persévérer, faire de son mieux, prendre au sérieux; prendre appui; enfoncer, fixer solidement BA⁴; centrer, concentrer; tendre, déployer, hisser; effort tendu KT, attention soutenue, sérieux⁵.

Cette liste des équivalents est suivie de seize exemples illustratifs.

Dans le dictionnaire Mandingue-Anglais (Vydrine 1999)⁶ qui traite du bambara et du maninka de Guinée, trois formes de ce lexème sont mentionnées: *bánban* (pour les deux langues en question), *bánba* (bambara) et *báman* (un dialecte bambara non-spécifié).

La polysémie du verbe est présentée en hiérarchie numérotée, chaque sens étant doté de l'indice concernant la structure des valences (transitif, intransitif, réfléchi ; là où d'autres valences sont attestées, la postposition introduisant l'oblique est indiquée). Pour les sens dont la répartition est limitée, des marques dialectales sont données : (m) pour maninka, (b) pour bambara. Voici les sens de *bánban* selon (Vydrine 1999), en notant que les expressions idiomatiques et les exemples illustratifs sont omis. Les traductions françaises des équivalents anglais sont données entre les crochets.

1. (m) *vt* stretch out tightly [tendre]
- 2.1. *vt* fix tightly; hammer in [fixer fortement ; enfoncer, clouer]
vr establish oneself firmly [s'établir solidement]
- 2.2. (m) *vt* strengthen [renforcer]
3. *vt* set sth. (against — *lá*), lean sth. (on, against — *lá*) [appuyer qqch. (sur — *lá*)), mettre (contre — *lá*)]
vr lean (on — *lá*, *kàn*, *mà*) [s'appuyer (contre — *lá*, *kàn*, *mà*)]
4. *vr* take pains (over — *lá*), be zealous (in — inf.) [faire un effort (pour — *lá*), faire du zèle (en — inf.)]
5. (m) *vt* support, give assistance [soutenir, donner l'assistance]
6. (bSegu) *vr* get off cheap, extricate oneself [se débrouiller, se tirer d'affaire]

Il s'avère que les deux dictionnaires se complètent :

— on trouve dans DUM les sens 'centrer, concentrer', 'hisser', 'prendre au sérieux', absents de VYD ;

⁴ Les index à deux lettres renvoient aux sources dont la liste est donnée dans l'Introduction du dictionnaire.

⁵ Les trois derniers sens sont nominaux. Dumestre considère les cas de conversion « verbe → nom » comme polysémie. Dans ce qui suit, les sens nominaux ne sont pas pris en considération.

⁶ Dans ce qui suit, VYD. Dans le dictionnaire de Charles Bailleul (Bailleul 2007) la polysémie de ce verbe est représentée plus laconiquement ; je ne la considère donc pas spécialement.

— on trouve dans VYD les sens ‘*vr* s'établir fermement’, ‘s'appuyer’, ‘se débrouiller, se tirer d'affaire’ (en dialecte de Ségou), absents de DUM. On y trouve également les sens *maninka* (mais non pas *bambara*) ‘*vt* renforcer’ (pourtant, on peut réunir ce sens avec 2.1 ‘fixer’) et ‘*vt* soutenir, donner assistance’ (en fait, on peut rapprocher ou même réunir ce sens avec 2.2. ‘renforcer’).

Il faut mentionner également que le sens 1 en VYD, ‘*vt* tendre’, marqué comme typique du *maninka*, apparaît dans DUM ; on peut donc conclure qu'il est typique de l'ensemble de l'aire *maninka* et *bambara*.

La polysémie consolidée de *bánban* peut être présentée comme suit :

1.1 *vt* tendre

1.2 *vt* hisser, déployer

2.1. *vt* fixer fortement ; enfoncer, clouer ; établir solidement
vr s'établir solidement

2.2. (m) *vt* renforcer

2.3. (m) *vt* soutenir, assister

3. *vt* appuyer qqch. (sur — *lá*), presser qqch. (contre — *lá*)
vr s'appuyer (contre — *lá*, *kàn*, *mà*)

4.1 *vr* faire un effort (pour — *lá*), faire du zèle (en — inf.)

4.2 *vr* prendre au sérieux

4.3 *vt* centrer, concentrer (sur — *kàn*)

5. (bSegu) *vr* se débrouiller, se tirer d'affaire.

2.2. *Bánban* dans le Corpus Bambara de Référence

En mai 2013, j'ai effectué une étude de la polysémie de ce verbe sur la base du Corpus Bambara de Référence qui comportait à l'époque environ 1,5 million de mots. J'ai trouvé 123 occurrences de *banban*, dont 12 (des emplois nominaux et des homonymes) ont été mis de côté, ce qui m'a laissé 111 occurrences (y compris quelques formes dérivées dont les sens sont en correspondance régulière avec les sens du verbe, et quelques emplois nominalisés du verbe dans les noms composés).

A ce moment-là, presque une moitié du Corpus (environ 700 000 du nombre total de 1,5 millions mots) était constitué par les textes de l'Ancien Testament. Le premier résultat intéressant a été que parmi les 111 occurrences de *bánban*, 76 (68,5%) provenaient de l'Ancien Testament. Autrement dit, il s'est avéré que dans cette source, *bánban* a été 2,5 fois plus fréquent (109 occurrences par million) que dans les autres textes du Corpus (42,5 occurrences par million).

On peut indiquer également que la variante phonétique *báman* (mentionnée dans les deux dictionnaires de référence, DUM et VYD) a eu 0 occurrences.

En septembre 2017, j'ai réanalysé l'utilisation de *bánban* dans le Corpus Bambara de Référence qui comptait à ce moment-là 4,1 million de mots en dépassant la version du mai 2013 de 2,73 fois. Conséquemment, le taux des textes de l'Ancien Testament s'était abaissé dans le Corpus à 17%.

Le nombre total des occurrences de *banban* (y compris les formes dérivées et les mots composés) a été de 253. Après la mise de côté des homonymes, des lexèmes dérivés irréguliers (comme le causatif *lábanban*) et des emplois nominaux, ce nombre a baissé à 226. Concernant le taux d'occurrences dans l'Ancien Testament (109 par million) et dans le reste du Corpus (44 par million), la tendance reste la même qu'en 2013.⁷

Malgré l'augmentation de l'échantillon, la variante phonétique *baman* n'est toujours pas apparue (plus précisément, toutes les 5 occurrences de la forme *bamanna* qu'on pourrait croire d'être des formes perfectives intransitives de notre verbe, se sont avérées des écritures erronées du mot *bámàna* 'bambara'). A l'évidence, cette variante n'appartient à l'usage écrit du bambara moderne.

Dans ce qui suit, je présenterai une analyse détaillée de la polysémie du verbe *bánban* sur la base des données du Corpus Bambara de Référence (version du septembre 2017). Dans cette analyse, la séquence des sens correspond, dans ses grandes lignes, à celle qu'on a vue plus haut. Les nouveaux sens seront introduits à côté de ceux qui leur sont les plus proches sémantiquement. Auprès de chaque sens, après la marque de valence (*vt*, *vi*, *vr*), le nombre d'occurrences du verbe *bánban* sera indiqué entre parenthèses.

Concernant ces données numériques, il faut préciser qu'à part les occurrences purement verbales, j'ai pris en compte les formes dérivées régulières et les mots composés où le sens verbal se manifestait clairement ; pour certains sens, le taux de fréquence de ces mots dérivés et composés s'est avéré assez élevé. Mais pour les sens où se manifeste la labilité verbale (ambivalence), l'opposition « transitif / intransitif / réfléchi », la nominalisation neutralise cette opposition (ex., 'établir solidement' vs. 's'établir solidement'), ce qui crée un problème pour les résultats statistiques. Un autre problème concerne les emplois passifs : faut-il les compter avec des occurrences intransitives ou transitives ? (J'ai pris la décision de les rajouter aux emplois transitifs : 'être établi' est compté avec 'établir solidement', plutôt qu'avec 's'établir solidement').

⁷ Comme je l'ai déjà dit, entre septembre 2017 et avril 2020 le Corpus a considérablement grandi et a atteint 11 140 640 mots. J'ai décidé de ne pas reprendre l'analyse du verbe *bánban* sur le nouvel échantillon : peut-être, cela permettrait de rajouter quelques nuances, mais je ne pense pas que cela modifierait cardinalement les choses. Cependant, j'utilise les données du Corpus-2020 là où cela semble nécessaire.

D'autre part, les calculs sont compliqués par le syncrétisme des sens. Il est souvent difficile de dire auquel des sens adjacents il faut attribuer un emploi du verbe, ex. : 2.1. 'établir solidement' ou 2.2 'renforcer' ? 4.1 'faire un effort pour' ou 4.2 'prendre au sérieux' ?

Il est vrai que ces réserves rendent approximative la distribution des occurrences par les sens. Et pourtant, il y a toutes les raisons d'affirmer que les données statistiques permettent d'identifier les tendances principales de la polysémie.

Considérons maintenant les sens du verbe *bánban* un par un.

1.1 vt (3) tendre

- (1) *Ò lá, áw bε jɔ̀ bánban jí` céfê...*
ce à 2PL.EMPH IPFV filet\ART tendre eau\ART parmi
'Alors vous tendez le filet au milieu du fleuve...' [Maléfices].

Dans VYD, ce sens est doté de la marqué dialectale (m) (= maninka), cependant, il apparaît dans DUM. Parmi les 3 occurrences, l'un provient du journal *Jekabaara*, où le verbe décrit des relations entre les gens (il s'agit donc de son emploi figuré), deux autres viennent de « Maléfices et manigances » et représentent donc la zone dialectale de Ségou. On peut donc enlever la marque dialectale, cependant, ce sens reste relativement rare en bambara.

1.2 vt (0) hisser (voile, drapeau)

Ce sens n'est pas représenté dans le Corpus où *bánban* ne se conjugue pas avec le mot *dàrapo* 'drapeau' ; d'ailleurs, on ne trouve pas ce sens parmi les exemples illustratifs dans DUM. En fait, pour 'hisser le drapeau', on utilise en bambara le verbe *yèlen*, et l'idée de « hisser les voiles » est étrangère aux Bambara qui sont un peuple foncièrement terrestre. Sans doute, pour *bánban*, il s'agit d'un sens marginal ou même fantôme ; probablement, on peut le réunir avec le sens 1.1 'tendre' ou l'éliminer carrément.

2.1. vt (22) fixer fortement ; enfoncer, clouer ; établir solidement

vi (3) s'établir solidement

vr (0) s'établir solidement

Il n'y a aucun exemple de réalisation de ce sens dans une construction réfléchie ; d'autre part, on trouve quelques cas d'emploi de construction intransitive, ce qui n'est pas prévu par les dictionnaires. Mais parmi les trois occurrences, un seul est incontestable, et dans les deux autres cas nous avons un participe résultatif *bánbannen* utilisé dans les contextes identiques (2). Il s'avère donc que l'emploi intransitif est marginal aussi.

(2) *Dùgukolo` bánban-nen bé kónùman àní à tɛ sé*
 terre\ART fixer-PTCP.RES être bien et 3SG IPFV.NEG arriver
kà lámàga.
 INF remuer

‘La terre est établie solidement, on ne peut pas la remuer’ [Layidu Koro].

Curieusement, parmi les 22 occurrences transitifs de *bánban* dans ce sens, 19 viennent de l'Ancien Testament (en s'agissant presque toujours de la Création), 2 viennent du journal *Jekabaara*, et un, de *Kibaru*. Il s'avère donc que le taux élevé des occurrences pour ce sens (environ 10% de toutes les occurrences du verbe) s'explique par « l'effet de buccin » décrit par Adam Kilgarriff (1997: 138–139) ;⁸ il n'est attesté que rarement en dehors de l'Ancien Testament.

2.2. vt (55) renforcer, rendre solide

vi (31) se renforcer, s'enraciner ; devenir stable

vr (3) se renforcer, s'enraciner

(3) *Nê míiri` lá, ò dè ka ni, ò be mɔ̀gɔ̀*
 1SG.EMPH pensée\ART à ce FOC QUAL bon ce IPFV personne\ART
bánban
 tendre

‘A mon avis, c'est ça qui est bon, cela renforce l'homme’ [Ce sigidamaw].

Dans VYD, le sens ‘renforcer’ est doté de la note (m) (= maninka), et il n'est pas mentionné dans DUM ; on pourrait donc penser qu'il n'existe pas en bambara. Et pourtant, ce sens s'avère le plus fréquent dans le Corpus Bambara de Référence, surtout l'interprétation abstraite de ‘renforcer’. En plus, des occurrences des emplois intransitifs (très nombreuses) et réfléchies (plus rares), non-attestés dans VYD, ont été trouvées.

La distribution des occurrences par sources est relativement équilibrée. Plus précisément, on y trouve aussi un déséquilibre au profit de l'Ancien Testament, mais à moindre degré que pour le sens 2.1.

Tableau 1. La distribution par les sources des occurrences de *bánban* en sens 2.2

	Ancien Testament	Jekabaara	Kibaru	autres
<i>vt</i>	25	21	4	5
<i>vi</i>	7	24	0	0
<i>vr</i>	2	0	1	0

⁸ En anglais, «the whelks problem». Il s'agit d'une situation où un échantillon comporte un texte traitant spécialement d'un objet (par exemple, le buccin, *whelks*) qui est assez rarement mentionné ailleurs. La présence de ce document rend la fréquence du mot en question exagérée et fausse la statistique.

2.3 vt (3) soutenir, assister

- (4) (*Ála...*) ò mîn ye nê bánban ka nê sìgi
 Dieu ce REL PFV.TR 1SG.EMPH tendre INF 1SG.EMPH asseoir
 nê fà` Dawidi ka màsa-sigilan` kàn
 1SG.EMPH père\ART Davide POSS roi-siège\ART sur
 ‘(Dieu...) qui m'a aidé et m'a fait asseoir sur le trône de mon père David’ [Layidu Kɔɔɔ].

Ce sens est absent dans DUM, et dans VYD il est marqué comme appartenant au maninka. Dans le Corpus, on le trouve dans l'Ancien Testament, d'où viennent les trois occurrences.

3. vt (1) appuyer qqch. (sur — *lá*), presser qqch. (contre — *lá*)

vr (0) s'appuyer (contre — *lá*)

vr (0) s'appuyer (contre — *mà*)

vr (22) s'appuyer (sur — *kàn*)

vi (12) s'appuyer (sur — *kàn*)

- (5) *Á* y' í jòngiri kà màrifa` bánban, kà múgu`
 3SG PFV.TR REFL s'agenouiller INF fusil-ART tendre INF poudre-ART
cì *dáje* *kélen ná.*
 frapper hyppotrague un à

‘Il s'est agenouillé, il a appuyé la crosse de son fusil (contre son épaule) et tiré sur un hippotrague’ [Bamanankan ni tubabukan].

Ce sens en emploi transitif est attesté dans VYD, où une expression est mentionnée : *kà màrifa` bánban* ‘appuyer la crosse du fusil (contre l'épaule, pour tirer)’.⁹

Aucune occurrence n'est attestée dans la construction réfléchie avec la postposition *lá*. Cependant, un exemple de cet emploi est donné dans DUM:

- (6) *Á* y' í bánban à *térike` lá* kà jòyɔɔɔ` sòɔɔ.
 3SG PFV.TR REFL tendre 3SG ami-ART à INF place-ART trouver
 ‘Il s'est appuyé sur son ami pour obtenir une place.’

La construction réfléchie avec la postposition *mà*, semble-t-il, n'est possible qu'avec le nom *dùgu* ‘terre’ (en bambara, la collocation *dùgu mà* est lexicalisée). Elle a été absente du Corpus en 2017, mais une occurrence est apparue dans la version du Corpus d'avril 2020. Par contre, la construction réfléchie avec la postposition *kàn* s'avère très fréquente, de même que la construction intransitive avec la même

⁹ Cependant, dans VYD, on trouve une autre expression: *kà màrifa` bánban à lá* ‘braquer le fusil sur qqn.’ Il n'est pas clair lequel parmi ces deux sens se réalise dans l'exemple (6).

postposition (non-attestée dans VYD ni DUM). La différence entre l'emploi réfléchi et intransitif est en conformité avec la tendance générale : avec le sujet animé, le verbe tend à être réfléchi, et avec le sujet inanimé, intransitif (Vydrine 1995: 171).¹⁰

L'analyse des occurrences (intransitives et réfléchies) avec la postposition *kàn* montre que dans presque tous les cas (à une exception près) il s'agit de 's'appuyer' dans le sens abstrait ('s'appuyer sur qqn. dans ses actions', 's'appuyer sur les règles', etc.), cf. (7). Par contre, dans une construction avec la postposition *lá*, *bánban* exprime le fait de s'appuyer physiquement, cf. (6).

- (7) *Û y' ù bánban nìn fɔ̃-li` kàn.*
 3PL PFV.TR 3PL tendre ce dire-NMLZ-ART sur
 'Ils se sont appuyés sur ces paroles' [Jekabaara 266].

Quant à la distribution de ces occurrences, un seul exemple vient du journal *Kibaru*, et 4 viennent de l'Ancien Testament. Tous les autres proviennent de *Jekabaara*,¹¹ et presque toutes appartiennent à la plume d'un seul auteur, Toumani Yalam Sidibé, le rédacteur-en-chef et auteur principal de ce journal. On peut conclure que l'emploi fréquent du verbe *bánban* dans ce sens est une particularité du style personnel de Toumani Yalam Sidibé.

4.1 vr (20) faire un effort (pour — *kàn*, dans — *fê*, pour — Inf., Sbjv), faire du zèle (en — *fê*, *kàn*)

Contrairement à VYD, l'oblique (l'objet d'application des efforts) est introduit par les postpositions *fê* (s'il s'agit d'une sphère d'activité, (8)) ou *kàn* (s'il s'agit du but d'activité, (9)), plutôt que *lá*.

- (8) *Án ka kán k' án bánban kóori-sene ìn fê...*
 1PL QUAL.AFF égal INF 1PL tendre coton-cultiver DEF par
 'Nous devons cultiver le coton avec du zèle...' [Jekabaara 116].

- (9) *Án ka dá án yèrê lá ; k' án bánban án ka*
 1PL SBJV coucher 1PL même à INF 1PL tendre 1PL POSS
ké-ta-w ké-cogo juman` kàn.
 faire-PTCP.POT-PL faire-façon bon\ART sur

'Nous devons avoir confiance en nous-mêmes et faire un effort pour faire bien tout ce qui doit être fait' [Jekabaara 113].

Assez souvent, le verbe apparaît en ce sens sans oblique.

¹⁰ Plus de la moitié des occurrences pour ce sens sont représentées par des mots composés ou dérivés, et le caractère réfléchi ou intransitif dans ces cas est hypothétique.

¹¹ En septembre 2017, les volumes de textes provenant de chaque journal, *Jekabaara* et *Kibaru*, étaient sensiblement égaux.

Concernant la distribution des occurrences de *bánban* en ce sens, *Jekabaara* prédomine (13 occurrences sur 20, dont 9 appartiennent à Toumani Yalam Sidibé), les autres apparaissent dans les sources diverses.

4.2 *vr* (14) **prendre au sérieux (qqch. — *kàn*, *mà*)**

Ce sens est mentionné dans DUM, il est absent de VYD. L'analyse du Corpus Bambara de Référence témoigne que le sens ‘prendre au sérieux’ appartient au noyau sémantique du verbe *bánban*. La postposition introduisant l'oblique (« ce qui est pris au sérieux ») n'est pas donnée dans DUM, mais les données du Corpus montrent que cette postposition est *kàn*, plus rarement *mà* (la nuance de sens entre ces deux postpositions dans ce contexte n'est pas claire).

- (10) *Án k' án bánban báara` kàn.*
 1PL SBJV 1PL tendre travail\ART sur
 ‘Prenons le travail au sérieux’ [Jekabaara 249].

Si on analyse la distribution des occurrences, *Jekabaara* (11 occurrences) prédomine toujours.

4.3 *vt* (11) **centrer, concentrer (sur — *kàn*)**

vr (6) **se concentrer (sur — *kàn*)**

- (11) *Ánw ODIMO m̀g̀w k̀ni ké-len b́ k' án*
 1PL.EMPH ABBR. personne-PL mais faire-PTCP.RES être INF 1PL
bárika bánban ò d̀ k̀n.
 force\ART tendre ce FOC sur

‘Mais nous, les gens de l'ODIMO, nous concentrons nos efforts sur cela’ [Jekabaara 91].

Ce sens aussi existe dans DUM, mais non pas en VYD. Il faut dire qu'il est parfois difficile de le distinguer (cela concerne surtout l'emploi réfléchi) des sens 4.1 et 4.2.

Toutes les occurrences réfléchies et 10 (sur 11) occurrences transitives proviennent de *Jekabaara*, dans la plupart de cas des écrits de Toumani Yalam Sidibé.

4.4. *vr* (9) **persister (dans — *kàn*), insister (sur — *kàn*)**

- (12) *Ù bɛ ù bánban nánbara` k̀n, ù bɛ ù b̀n*
 3PL IPFV 3PL tendre fraude\ART sur 3PL IPFV 3PL refuser
k̀ k̀s̀gin.
 INF revenir

‘Ils persistent dans l'injustice, ils refusent de revenir’ [Layidu K̀r̀].

Ce sens n'apparaît ni dans DUM, ni dans VYD. Sa proximité sémantique avec les sens 4.1, 4.2, 4.3 est indubitable, mais il est pourtant nettement à part. 5 occurrences

de ce sens proviennent de *Jekabaara* (dont une seule appartient à T. Ya. Sidibé), les autres 4 viennent de l'Ancien Testament.

5. (*bSegu*) *vr* (0) se débrouiller, se tirer d'affaire

Ce sens est mentionné dans VYD (sans exemples illustratifs), les données du Corpus ne le confirment pas.

On trouve dans le Corpus deux autres sens pour lesquels je n'ai trouvé qu'un seul exemple dans l'échantillon de septembre 2017 (dans les deux cas, la source est *Jekabaara*, et l'auteur est T. Ya. Sidibé).

6. *vi* (1) peser lourdement (sur — *kàn*)

(13) *Sarili Tayilɔri ka fànga` gèleya` bánban-na*
Charles Taylor POSS force\ART lourdeur\ART tendre-PFV.INTR

kósebe liberiya-ka-w kàn...

très Libérien-GENT-PL sur

‘Le pouvoir de Charles Taylor pesait lourd sur les Libériens...’
[jekabaara281_8sidibe_wale].

7. *vi* (1) avoir pour cause (qqch. — *lá*), provenir (de — *lá*)

(14) *Fóyì wére ma ké jàmana-den-w ka híne`*
rien autre PFV.NEG faire pays-enfant-PL POSS pitié\ART

bó-jɔɔn-na` kó, n' ò bánban-na jàmana`
sortir-RECP-à\ART après si cela tendre-PFV.INTR pays\ART

jé-maa-w ka jò-yɔrɔ` fá-bali-ya` lá
œil-personne-PL POSS dresser-place\ART remplir-PTCP.PRIV-ABSTR\ART à

kà táa sùgu-kɔnɔ-fen-w sòngɔ` dàn-sigi-li` mà.
INF aller marché-dans-chose-PL prix\ART limite-asseoir-NMLZ\ART ADR

‘Rien d'autre n'a allégé le lot des habitants du pays, et la cause en est que les dirigeants du pays n'ont pas fait ce qu'ils devraient faire pour limiter les prix des marchandises aux marchés’ [Jekabaara 276].

Pour 5 occurrences (moins de 2% de l'échantillon) le contexte est insuffisant pour définir le sens.

2.3. Un résumé pour *bánban*

L'analyse des occurrences du lexème dans le Corpus Bambara de Référence permet de considérablement préciser et détailler la présentation de ses caractéristiques sémantiques, grammaticales et stylistiques dans le dictionnaire :

— détecter les sens qui n'avaient pas été répertoriés par des dictionnaires « classiques » (2.2. ‘renforcer’ ; 2.3. ‘assister’ ; 4.4 ‘persister, insister’; 6. ‘peser lourdement’ ; 7. ‘provenir’) et, éventuellement, de se débarrasser des sens fantômes ;

— préciser l'information concernant les collocations et les valences verbales ;
— ranger les sens par fréquence. Ainsi, pour *bánban* le sens ‘tendre’ (le fil de trame; une corde, un filet de pêche...) qui peut être considéré comme étymologiquement primaire (et pour cette raison il est mis en premier lieu dans VYD) s'avère peu fréquent dans le bambara moderne. Si on veut rendre le dictionnaire plus commode aux utilisateurs, il serait probablement plus approprié de donner les premières places aux sens avec les indices statistiques les plus élevés (ou, au moins, les données statistiques pourraient être présentées pour mieux orienter les utilisateurs) ;
— prendre en compte les particularités des sources. Ainsi, il s'est avéré que les sens 2.1, 2.2, 2.3 de *bánban* (d'ailleurs, avec de bons indicateurs statistiques) sont typiques surtout des textes de l'Ancien Testament. D'ailleurs, la plus grande partie des occurrences de ce verbe vient soit de l'Ancien Testament, soit des textes écrits par Toumani Yalam Sidibé (originaire de Kita). Évidemment, cette conclusion est d'intérêt pour la lexicographie bambara.

3. Jàhadi

3.1. *Jàhadi* dans les dictionnaires « classiques »

Ce mot est un emprunt à l'arabe où *jihad* signifie ‘effort’ (surtout un effort pour la religion) ; ‘lutte’ ; ‘guerre’ (religieuse).

Dans le dictionnaire (Vydrine 1999), l'entrée *jàhadi* est présentée d'une façon préliminaire, avec deux sens:

1 *jihad* (*Muslim holy war*) 2 (m) calamity

Dans (Bailleul 2007) aussi, *jàhadi* a deux sens:

1 grand malheur, catastrophe, accident grave
2 guerre ‘sainte’ (pour les musulmans)

Enfin, (Dumestre 2011), qui mentionne deux variantes phonétiques de ce mot, *jàhadi* et *jàadi*, donne les sens suivants :

catastrophe, malheur ; guerre KT, guerre sainte

Les trois dictionnaires s'accordent donc en ce qui concerne les sens ‘catastrophe, malheur, calamité’ et ‘guerre sainte’ (pour les musulmans), et DUM y rajoute le sens ‘guerre’.

3.2. *Jàhadi* dans le Corpus Bambara de Référence

Dans le Corpus Bambara de Référence consulté en avril 2020, j'ai trouvé 161 occurrences de ce lexème. Elle apparaît sous trois variantes phonétiques : *jàhadi* (133), *jàadi* (12) et *jihadi* (16), dont la dernière est absente des dictionnaires « classiques ». ¹²

Quant à la polysémie de ce lexème, elle s'est avérée beaucoup plus riche dans le Corpus que dans les dictionnaires. Dans ce qui suit, chaque sens détecté sera présenté et discuté.

I calamité, malheur ; catastrophe (75 occurrences)

(15) *Kèlè` , jànkó múgu` ni kìsè` mánà dòn à*
 guerre\ART surtout poudre\ART et grain\ART COND.AFF entrer 3SG
lá, à bε ké kó cèjugu yé, à bε ké jàhadi yé.
 à 3SG IPFV faire affaire vilain PP 3SG IPFV faire calamité PP

‘La guerre, surtout si on met en action des armes à feu, c'est une chose vilaine, cela devient une calamité’ [Kibaru 24].

Ce sens est de loin le plus fréquent, ce qui confirme l'intuition de Dumestre et Bailleul qui l'ont mis en première position dans leurs dictionnaires.

J'ai d'abord attribué le sens ‘guerre’ (mentionné dans DUM) à 4 occurrences. Cependant, après réflexion, je suis arrivé à la conclusion que dans un cas (18), le sens peut être interprété comme ‘troubles’, et dans les 3 autres cas, comme ‘calamité’, en particulier dans l'expression *jàhadi kèlè* (16)¹³ ou dans le contexte suivant : *Ségu Damonson ka jàhadi tuma`* [Jεemu ni jiidi] qui signifie, sans doute, ‘la période néfaste de Da Monson de Ségou’. Les données du Corpus ne confirment donc pas l'existence du sens ‘guerre’ du mot *jàhadi*.

(16) *Fabali nà-na Nataala, à ma mén, shúbaa-w ni*
 Fabali venir-PFV.INTR Nataala 3SG PFV.NEG durer sorcier-PL et
nyàgwan-w ye shí ban kèlè` dábɔ,
 sorcier.bienfaisant-PL PFV.TR semence finir guerre\ART commencer
jàadi kèlè-ba.
 calamité guerre-AUGM

‘Quand Fabali est venu à Nataala, peu après, les sorciers et les contre-sorciers ont entamé une guerre d'extermination, une guerre dévastatrice’ [Jama 5].

¹² Dans le dictionnaire Bamadaba, <http://cormand.huma-num.fr/bamadaba.html>, outre ces trois variantes, on trouve une quatrième, *jàhadu*, mais cette forme n'apparaît pas dans le Corpus.

¹³ Dans d'autres contextes, *jàhadi kèlè* signifie ‘une guerre sainte (des musulmans)’, cf. plus bas ; mais dans la phrase (16) il ne peut pas s'agir d'une guerre pour l'Islam.

Curieusement, on ne trouve qu'une seule occurrence de la variante *jìhadi* avec ce sens (contre 63 *jàhadi* et 8 *jàadi*).

2.1 troubles (15)

Dans la grande majorité de cas, ces occurrences se réfèrent aux événements de mars 1991 (la chute du régime de Moussa Traoré et les troubles qui l'ont accompagnée), comme dans (17). On trouve quand même quelques cas où il s'agit d'autres troubles, mais toujours postérieurs à mars 1991 (18). On peut donc supposer que ce sens du mot *jàhadi* a émergé en mars 1991. Il n'apparaît que dans la presse, aux articles concernant les actualités politiques.

(17) *Í n' à fɔ bée b' à kàlamà sà̀n 1991*
 2SG CERT 3SG dire tous IPFV 3SG au.courant.de an 1991
márisi-kalo jàhadi` nà-na ni tákisi-ko` ni
 mars-mois calamité\ART venir-PFV.INTR et taxe-affaire\ART et
wúsuru-ko` láafu-li` yé.
 patente-affaire\ART avilir-NMLZ\ART PP

'Comme tout le monde sait, les troubles du mois de mars 1991 ont entraîné de la négligence par rapport aux taxes et impôts' [Kibaru 359].

(18) *Sègin` tùn ka kán kà ké ntòlatan in kàn dón`*
 retour\ART PST QUAL.AFF égal INF faire football ce sur jour\ART
mîn Bàmakɔ, ò bó-ra à mà, k' à dá à kàn
 REL Bamako ce sortir-PFV.INTR 3SG ADR INF 3SG poser 3SG sur
jàhadi` tùn wúli-la Álizeri, m̀̀gɔ cáman nín`
 calamité PST se.lever-PFV.INTR Alger personne nombreux âme\ART
tó-ra mîn ná.
 rester-PFV.INTR REL à

'Le jour où ce match devait avoir lieu de nouveau à Bamako, il a été ajourné, car des troubles ont commencé à Alger, dans lesquelles beaucoup de gens ont péri' [Jekabaara 35-36].

Parmi les 15 occurrences, la variante *jàadi* est apparue une fois, et *jàhadi*, 14 fois.

2.2 crime (7)

(19) *Á be sé kà ké fána jàhadi nìn-nu ké-baga`*
 3SG IPFV arriver INF faire aussi calamité ce-PL2 faire-AG.OCC\ART
te dón kà dá à kàn sàriya-w be ù jémadògo.
 IPFV.NEG connaître INF poser 3SG sur loi-PL IPFV 3PL dissimuler

'Il se peut aussi que les auteurs de ces crimes restent inconnus, parce que les lois les cachent' [Ntomokun].

Dans toutes les occurrences de ce sens, le lexème apparaît sous sa variante *jàhadi*.

3 guerre sainte (19)

Ce sens se réalise souvent dans le cadre de l'expression *jàhadi keɛ* (20), mais pas seulement (21).

(20) *Á fúruce` bàn-na sèn-na-kolon diine` mànton-w*
 3SG époux\ART refuser-PFV.INTR pied-à-nu religion\ART peuple-PL
ka jáhadi keɛ` lá.
 POSS calamité guerre\ART à

‘Son époux a refusé de participer à la guerre sainte des adeptes de la « religion des nu-pieds »¹⁴ [Kibaru 319].

(21) *Írisi-w bìn-nen` Áfuganisitan kàn k' à mìnɛ sà̀n*
 russe-PL tomber-PTCP.RES\ART Afganistan sur INF 3SG attraper an
1979, silame-w ye jáhadi` wúli írisi-w kámà.
 1979 musulman-PL PFV.TR calamité\ART se.lever russe-PL pour

‘Quand les Russes ont attaqué l'Afganistan et l'ont occupé en 1979, les musulmans ont commencé une guerre sainte contre les Russes’ [Kibaru 472].

La majorité des occurrences pour ce sens de *jàhadi* vient des journaux *Kibaru* et *Jekabaara*, deux proviennent de livres sur l'histoire, et deux, du Coran.

Concernant la forme du mot, nous avons 1 occurrence de *jàadi* et 4 *jìhadi* (toutes les quatre dans un article de *Kibaru* No. 532), dans les autres 14 cas (y compris les deux venant du Coran), c'est *jàhadi*.

4 effort religieux (32)

(22) *Á yé jáhadi` ké ála-ko` lá, jáhadi lakika` lá*
 2PL IMP calamité\ART faire Dieu-affaire\ART à calamité vrai à
ála yé!
 Dieu PP

‘Faites un effort pour la cause de Dieu, un vrai effort pour Dieu !’ [Kurana 22].

Très souvent, ce sens est réalisé dans le contexte *kà jìhadi` ké Ála ka síra` kàn* ‘faire un effort sur le chemin de Dieu’.

La quasi-totalité des occurrences (plus précisément, 30) vient du Coran, les deux restantes apparaissent dans des textes de Hadiths ou des citations de Hadiths. Il semble que ce soit le seul sens de notre lexème admis dans le Coran (à part le deuxième sens religieux, ‘guerre sainte’, cf. ci-dessus) : on ne trouve pas dans le Livre Saint d'occurrences de *jàhadi* pour ‘calamité’, ‘trouble’ ou ‘crime’. Je suppose qu'Alihaji

¹⁴ Une secte musulmane active au Mali dans les années 1990.

Modibo Jara, le traducteur de Coran en bambara, a sciemment évité d'utiliser ce mot dans un sens profane.

Une autre particularité intéressante du sens en question concerne la variabilité phonétique. Parmi les 32 occurrences, nous trouvons 10 *jìhadi* (9 dans le Coran, 1 dans le livre de Hadiths). Il est vrai que la variante *jàhadi* reste prédominante, mais le ratio est 1:2, plutôt que 1:8 pour la totalité d'échantillon (et si on compte le taux pour toutes les occurrences du lexème dans des sens autres que 'effort religieux', on trouve presque 1 *jìhadi* pour 50 *jàhadi* + *jàadi*). Évidemment, nous observons une tendance à utiliser la forme phonétiquement proche de l'original arabe pour le sens « canonique », sans doute, pour éviter des confusions avec les sens « profanes ». Par contre, dans l'Ancien Testament nous ne trouvons que la forme *jàhadi* avec le sens profane 'calamité, désastre'.

5. objet magique, fétiche (6)

Toutes les occurrences (dans la forme *jàhadi*) viennent d'une seule source, l'interview de Souleyman Kantè par Amadou Tamba Doumbia, et se réfèrent à un seul objet, le fétiche qui assurait la puissance de Sumanguru Kantè, le roi de Soso.

(23) *Bá` dè n' à ka jàhadi` bó-ra*
mère\ART FOC et 3SG POSS calamité\ART sortir-PFV.INTR
jíne-w bólo.
esprit-PL CNTRL

'C'est la mère qui a apporté son fétiche de chez les esprits'.

Il reste 6 occurrences où le contexte est insuffisant pour identifier le sens. Cela constitue moins de 5% du total et ne doit donc pas influencer notablement les conclusions.

3.3. Un résumé pour *jàhadi*

Comme ce mot est emprunté à l'arabe, il est certain que ses sens primaires soit 'effort (religieux)' et 'guerre sainte, guerre de religion'. Logiquement, on peut supposer une direction suivante de l'évolution sémantique pour ce lexème:

(en arabe : effort religieux > guerre sainte) > calamité, catastrophe, malheur > crime > troubles

Il s'avère qu'en bambara les nouveaux sens (surtout 'calamité, catastrophe, malheur') ont pris le dessus en éclipsant le sens original ('effort religieux'), qui reste réservé aux contextes religieux.

Le fait que la variante dominante, *jàhadi*, manifeste une modification de la voyelle par rapport à la forme arabe, de même que la prépondérance du sens 'calamité' (non-existant en arabe) témoigne au profit du caractère ancien de cet emprunt. Par

conséquent, la variante *jihadi*, sans doute, résulte d'un emprunt plus récent, ce qui explique son utilisation presque exclusive dans le sens original ('effort religieux').¹⁵

Les deux petites études que je viens de présenter visent à démontrer le potentiel du Corpus Bambara de Référence pour la lexicographie « classique ». ¹⁶ Les avantages du Corpus semblent évidents, et il serait difficile de mettre en question son indispensabilité dans l'élaboration des dictionnaires académiques bambara à l'avenir. Cependant, je vois ici au moins deux questions :

— est-ce que le Corpus Bambara de Référence est suffisamment grand et équilibré pour qu'un dictionnaire dérivé de ce Corpus représente la richesse lexicale de la langue d'une façon adéquate ?

En avril 2020, ce Corpus comptait plus de 11 millions de mots (dont plus de 1,5 million dans le sous-corpus désambiguïsé). Il comportait une grande majorité de tous les textes publiés en bambara jusqu'ici,¹⁷ et si on le juge déséquilibré, cela veut dire que l'ensemble de la production écrite bambara est déséquilibré aussi. On peut se rappeler que le premier dictionnaire dérivé d'un corpus annoté, Cobuild English Dictionary (1987), se basait sur un corpus de 6 millions mots (puis revérifié sur un corpus de 9 millions) ; le Corpus Bambara dépasse ce chiffre ;

— serait-il réaliste de créer un dictionnaire dont chaque entrée se baserait sur le concordancier complet (de la façon dont je l'ai essayé de faire ici pour ces deux lexèmes) ?

Cette question est plus épineuse : la compilation d'un dictionnaire dérivé d'un grand corpus s'avère très complexe, elle sous-entend une analyse profonde du concordancier

¹⁵ En maninka de Guinée, il n'existe qu'une seule forme, *jàhadi*.

On peut se poser une question : ne faut-il pas considérer *jàhadi* et *jìhadi* en bambara comme deux lexèmes différents (de même que, par exemple, *chef* et *chief* en Anglais) ? Cette solution est tout à fait envisageable, mais pour le moment, je les traiterai comme des variantes d'un seul lexème, car parmi les sens de *jìhadi* il n'y en a aucun qui ne serait pas valable pour *jàhadi*.

¹⁶ Je souligne qu'il s'agit des études lexicographiques plutôt « classiques », où le Corpus ne sert que comme source des exemples naturels (même si c'est une source très riche). L'application des méthodes du traitement automatique de langue (TAL) pour la lexicographie est un sujet à part, je n'y touche pas dans cet article.

¹⁷ D'après mon évaluation très approximative, entre 70 et 80%. Sans doute, certaines publications manquantes seront retrouvées, numérisées et intégrées dans le Corpus, mais une partie risque d'avoir disparu pour toujours. Mais il ne faut pas oublier que le bambara est une langue vivante, et que de nouveaux textes bambara sont générés sans cesse.

de chaque lexème. Les deux lexèmes analysés dans cet article, *bánban* et *jàhadi*, sont de fréquence moyenne.¹⁸ Le tri manuel des concordanciers de ces deux lexèmes m'a pris environ une journée de travail. Conséquemment, pour un dictionnaire de 12 000 à 15 000 entrées, on aurait besoin de dizaines années de travail d'un lexicographe qualifié, et il n'y a aucune garantie qu'une vie suffirait pour achever un tel travail.

Une solution qui s'impose pourrait être celle d'un projet collectif. D'ailleurs, des gros dictionnaires académiques sont le plus souvent créés par des collectifs de lexicographes.

Pour les lexèmes très fréquents, on peut penser à se limiter à des échantillons limités randomisés, ce qui pourrait donner une économie considérable.

Une autre piste pourrait être l'application de la méthode de la désambiguïsation automatique des sens basée sur la mise en grappe contextuelle. Il faut avouer que les résultats de premiers essais entrepris en 2018 ont été plutôt décevants, mais des modifications de l'algorithme pourraient probablement les améliorer. Cf. la publication (Maslinsky 2019) où certaines pistes allant dans ce sens ont été tracées.

Gloses et abréviations

1, 2, 3 — 1, 2, 3 personne

ABBR — abréviation

ADR — postposition à valeur adressative et directive

AFF — affirmatif

ART — article tonal

AUGM — suffixe augmentatif

(b) — bambara

(bSegu) — bambara de Ségou

CERT — futur certain

CNTRL — postposition avec la valeur de contrôle

COND — conditionnel

DEF — le « nouvel article défini »

EMPH — pronom personnel emphatique

¹⁸ Dans NoSketchEngine, le moteur de recherche du Corpus Bambara de Référence, il y a une option de génération de la liste des lexèmes en ordre décroissant de fréquences. Malheureusement, cette option n'est pratiquement pas utilisable pour le corpus entier (à cause de l'homonymie abondante), et même pour le sous-corpus désambiguïsé, l'ordre des fréquences est sensiblement perturbé pour plusieurs raisons. Cependant, on peut évaluer approximativement le rang de *bánban* dans cette liste autour du 1500ème, et celui de *jàhadi* autour du 2000ème.

- FOC — particule de focalisation
GENT — suffixe de nom d'habitant ou originaire d'une localité
IMP — impératif
INF — infinitif
INTR — intransitif
IPFV — imperfectif (inaccompli)
(m) — maninka
NEG — négation, négatif
NMLZ — suffixe de nom d'agent
PFV — perfectif (accompli)
PL — pluriel (-w)
PL2 — marque non-productive du pluriel (-lu)
POSS — marque possessive
PP — postposition au sémantisme large (yé)
PST — marque du passé discontinu
PTCP.POT — participe potentiel
PTCP.RES — participe résultatif/statif
QUAL — marque affirmative de l'énoncé qualitatif
RECP — pronom réciproque ; suffixoïde à valeur de « nom de partenaire »
REFL — pronom réfléchi
REL — relativisateur
SBJV — subjonctif
SG — singulier
TR — transitif
vi — emploi intransitif d'un verbe
vt — emploi transitif d'un verbe
vr — emploi réfléchi d'un verbe

Sources

- [Bamanankan ni tubabukan] — *Bamanankan ni tubabukan kalancogo juman. Méthodologie convergente de l'enseignement des langues nationales de du français*. Bamako: IPN - DNAFLA – DNEF, 1994, 28 p.
[Cε sigidamaw] — Diarra, Justin (Jōnke). *Étapes de la vie de l'homme Bambara dans la société traditionnelle (Cε sigidamaw)*. Édité par José Morales. Falaje.
[Jama] — *Jama* (une revue irrégulière).
[Jεemu ni jiidi] — Jara, Modibo. *Jεemu ni jiidi*. Bamako: Imprimerie SISA.
[Jεkabaara] — *Jεkabaara* (mensuel)
[Kibaru] — *Kibaru* (mensuel)

- [Kurana] — *Kuranε kɔrɔ bayɛɛmanen bamanankan na. Naaniye fɔlɔ*. Jara, Alihaji Modibo (traducteur). Bamako, 2011.
- [Layidu Kɔrɔ] — Layidu Kɔrɔ. Bamako, 2006.
- [Maléfices] — Dumestre, Gérard; Touré, Seydou. *Maléfices et manigances : Chroniques maliennes*. (version bambara).
- [Ntomokun] — Nare, Sanba. Ntomokun. Bamako: EDIS, 2010.

Bibliographie

- 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.humanum.fr/bamadaba.html>.
- Dumestre, Gérard. 2011. *Dictionnaire bambara-français suivi d'un index abrégé français-bambara*. Karthala. Paris.
- Hanks, Patrick. 2012. The corpus revolution in lexicography. *International Journal of Lexicography* 25(4). 398–436.
- Kilgarriff, Adam. 1997. Putting frequencies in the dictionary. *International journal of lexicography* 10(2). 135–155.
- Kilgarriff, Adam, Michael Rundell & Elaine Uí Dhonnchadha. 2006. Efficient corpus development for lexicography: building the New Corpus for Ireland. *Language Resources and Evaluation Journal* 40(2). 127–152. <https://doi.org/10.1007/s10579-006-9011-7>.
- Maslinsky, Kirill. 2019. Positional skipgrams for Bambara: a resource for corpus-based studies. *Mandenkan* 62. 165–183. <https://doi.org/10.4000/mandenkan.2119>.
- Vydrin, Valentin. 2013. Bamana Reference Corpus (BRC). In Chelo Vargas-Sierra (ed.), *Procedia - Social and Behavioral Sciences: Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013)*, vol. 95, 75–80. Alicante: Elsevier. <http://www.sciencedirect.com/science/journal/18770428>.
- Vydrin, Valentin. 2014. Projet des corpus écrits des langues manding : le bambara, le maninka. In Mathieu Mangeot, Fatiha Sadat & Brigitte Bigi (eds.), *Actes de l'atelier sur le traitement automatique des langues africaines TALAf 2014. Actes des Ateliers TALN 2014*. Marsel. <http://www.taln2014.org/site/actes-en-ligne/actes-en-ligne-ateliers/>.

- Vydrin, Valentin. 2018. *Malidaba : Dictionnaire électronique maninka-français-anglais-russe, avec des index français-maninka, anglais-maninka, russe-maninka*. <http://cormand.huma-num.fr/cormani/dictionnaire.html>.
- Vydrin, Valentin, Kirill Maslinsky & Jean-Jacques Méric. 2011. *Corpus Bambara de Référence*. <http://cormand.huma-num.fr/index.html>.
- Vydrin, Valentin, Andrij Rovenchak & Kirill Maslinsky. 2016. Maninka Reference Corpus: A presentation. In *TALAf 2016 : Traitement automatique des langues africaines (écrit et parole) Atelier JEP-TALN-RECITAL 2016*. Paris. http://talaf.imag.fr/2016/Actes/VYDRIN_ET_AL%20-%20Maninka%20Reference%20Corpus:%20A%20Presentation.pdf.
- Vydrine, Valentin. 1995. Verbes réfléchis bambara. Part 2: Classification des verbes réfléchis. *Mandenkan* 29. 1–114.
- Vydrine, Valentin. 1999. *Manding-English Dictionary (Maninka, Bamana)*. Vol. 1. St. Petersburg: Dmitry Bulanin Publishing House.

Valentin Vydrin

Towards corpus-driven dictionaries for Manding languages

Availability of big electronic annotated corpora for Maninka and Bamana languages lets us think about development of corpus-driven dictionaries, i.e. dictionaries of a new generation. In a corpus-driven dictionary, elaboration of the polysemy, collocations, syntactic valencies, stylistic, regional and dialectal characteristics of each lexeme (and of each sense of a lexeme) will be based on the analysis of a concordance extracted from the corpus. Henceforth, it becomes possible to evaluate statistically each feature. Two sample examples of lexicographic analysis of Bambara lexemes, *bánban* and *jàhadi*, are given.

Key words: Bamana language, Bambara language, Manding, corpus-driven lexicography, polysemy

Valentin Vydrin

Vers la lexicographie mandingue sur la base des gros corpus annotés

La disponibilité des corpus électroniques annotés comptant des millions de mots pour les langues maninka et bambara permet de penser des dictionnaires de nouvelle génération. Dans un tel dictionnaire, l'élaboration de la polysémie, combinabilité, valences syntaxiques, caractéristiques stylistiques, régionales et dialectales pour chaque lexème (et pour chaque sens du lexème) sera basée sur l'analyse du concordancier dérivé du corpus. Désormais, il devient possible d'évaluer statistiquement chaque trait caractéristique du lexème. Une analyse lexicographique,

sur la base des concordanciers, est présenté sur l'exemple des lexèmes bambara *bánban* et *jàhadi*.

Mots clé : bambara, mandingue, lexicographie dérivé du corpus, polysémie

В. Ф. Выдрин

Корпусная лексикография для языков манден

Создание миллионных электронных аннотированных корпусов для языков манинка и бамана ставит вопрос о разработке «корпусных словарей», т.е. словарей нового поколения. В таком словаре разработка полисемии, сочетаемости, синтаксических валентностей, стилистических и регионально-диалектных характеристик каждой лексемы (и каждого значения лексемы) должно базироваться на анализе корпусного конкорданса. Становится возможным не только выявление различных характеристик, но и их статистическая оценка. В статье даётся, в качестве примера, лексикографический анализ баманских лексем *bánban* и *jàhadi* на основе корпусных данных.

Ключевые слова: язык бамана, манден, корпусная лексикография, полисемия