Les Cahiers du CEDIMES

2022, Volume 17, Numéro 2, https://doi.org/10.69611/cahiers17-2-06

ISSN 2110-6045

L'EXPLOITATION DE L'ENVIRONNEMENT LEXICAL DANS LE TRAITEMENT AUTOMATIQUE DU KA POSSESSIF EN BAMANANKAN

Issiaka BALLO
Faculté des Lettres, des Langues, et des sciences du langage (FLSL)
Université des Lettres et des Sciences Humaines de Bamako (ULSHB), Bamako, Mali

issiakaballo79@gmail.com

Résumé :
Pour traiter le bamanankan (bambara) dans un programme d'intelligence artificielle, il faudra résoudre au préalable le problème des unités homographes de catégorie lexicale distincte. Le ka illustre bien ce phénomène de recoupement d'orthographe : ka (marqueur de l'adjectif qualificatif), ka (infinitif), ka (possession), ka (impératif). Il faut donc des indices matériels pour distinguer les ka entre eux dans un texte. À ce titre, la présente étude cherche à démasquer les indicateurs lexico-syntaxiques capables de désambiguïser les ka de la possession des autres ka se situant dans d'autres catégories. Les usages pratiques relevant du traitement artificiel du ka de la possession sont nombreux, étant entendu que ces techniques interviennent au moins dans la traduction automatique. Le présent travail formule des hypothèses sur les indicateurs lexicaux qui désambiguïsent le ka. Les hypothèses sont vérifiées à travers des termes de recherche appropriés qui extraient les occurrences respectives d'un corpus à l'aide du logiciel Bosolan. Les occurrences sont étudiées pour distinguer les cas de confirmation et d'infirmation de l'hypothèse. En effet, des indicateurs matériels pouvant guider un programme informatique ont été trouvés à travers l'exploitation de l'environnement lexical. Sur les 6 hypothèses formulées, plus de la moitié des occurrences a été confirmé.

***Mots-clés** : catégorie, corpus, environnement lexical, intelligence artificielle, occurrence.*

Abstract:
To process bamanankan (bambara) in an artificial intelligence program, it will first be necessary to solve the problem of homograph units of distinct lexical category. The ka is a good illustration of this spelling overlap phenomenon: ka (marker of the qualifying adjective), ka (infinitive), ka (possession), ka (imperative). It is therefore necessary to have material indices to distinguish the ka occurrences from each other in a text. As such, the present study seeks to unmask the lexico-syntactic indicators capable of disambiguating the ka of the possession from other ka with different categories. The potential uses of the artificial processing of possession ka are numerous, as these techniques are involved at least in machine translation. This work formulates hypotheses on lexical indicators that disambiguate the ka. The hypotheses are verified through appropriate search terms which extract the respective occurrences from a corpus using the Bosolan software. The occurrences are studied to distinguish between confirming and disproving cases of the hypothesis. Indeed, material indicators that can guide a computer program have been found through the exploitation of the lexical environment. About 6 stated hypotheses, more than the average occurrences were confirmed.

***Keywords** : category, corpus, lexical environment, artificial intelligence, occurrence.*

***Classification JEL :** Z0 .*

Introduction

Les langues africaines ont accusé du retard sur le plan de leur implémentation dans l'outil informatique. Ces retards vont du simple traitement de textes jusqu'aux astuces du *big data* et à la communication en ligne. Les chercheurs devraient s'en préoccuper autrement le fléau ne fera qu'aggraver le phénomène d'appauvrissement des langues africaines. En témoigne le cri du cœur de Camara (1996 : 812) : « L'internet risque d'accroître le phénomène d'appauvrissement linguistique en favorisant l'usage exclusif de l'anglais. La seule façon de contrecarrer cette tendance est de fournir aux langues minoritaires un support informatique adéquat, depuis la simple possibilité de transmission des caractères à travers les réseaux, jusqu'aux outils évolués ».

A l'instar de la plupart des langues africaines, le bamanankan (bambara) du Mali est menacé par cette carence quant à son traitement numérique. Il est de notoriété publique que le traitement artificiel, suppose l'implémentation d'un certain nombre d'outils sous forme de programmes informatiques. Ces outils vont des aides à la rédaction aux aides à la traduction en passant par les aides à l'enseignement des langues assisté par ordinateur. Cependant, que faut-il faire pour avoir une implémentation effective de ces outils en bamanankan ? Il s'agit essentiellement de faire un transfert de l'intelligence humaine à l'intelligence artificielle des stratégies de reconnaissance des unités lexicales d'un corpus. L'unité lexicale *ka* fait bien partie des cas d'implémentation en bamanankan. Or, pour implémenter la reconnaissance automatique des occurrences du *ka* de la possession, l'on se heurte à plusieurs obstacles. L'un des obstacles les plus saillants est le phénomène du recoupement d'orthographes de plusieurs *ka* au statut lexical distinct. En l'occurrence, pour qu'une machine reconnaisse le *ka* de la possession des autres *ka* potentiellement présents dans un corpus, il faut nécessairement passer par l'identification des stratégies qui permettent de désambiguïser le *ka* de la possession.

Dans cette perspective, l'objectif de la présente contribution est la désambiguïsation des *ka* de la possession à travers l'exploitation de l'environnement lexical. Cela représente une contribution significative dans le cadre de son traitement numérique.

Comme hypothèse de traitement, nous retenons que le *ka* de la possession peut être désambiguïsé suite à l'exploitation de l'environnement lexical des constituants de la phrase. Cette hypothèse sera testée par l'utilisation des moyens (voir §1) dont nous disposons pour l'extraction des phrases contenant l'occurrence *ka*.

La méthodologie utilisée pour établir le traitement artificiel du *ka* de la possession est détaillée dans les lignes qui suivent.

De prime abord, un corpus d'expressions bamanan « Ɲɛɲɛkɔrɔ *ka* Tonn*ka*n (ƝT) » a été choisi pour l'identification et le dépouillement des occurrences de *ka* qui font l'objet de ce travail. Il s'agit d'un corpus d'environ 50 000 occurrences. On y a dénombré 2 945 (2 280 *ka* et 665 k') occurrences totales de *ka* parmi lesquelles il y a eu 774 occurrences du *ka* de la possession. Cela fait un pourcentage de 1,54% des 50 000 occurrences totales du corpus et 26,28% des 2 945 *ka* du corpus.

Cependant, l'analyse des occurrences du *ka* de la possession est structurée en des paragraphes distincts. Le libellé de chaque paragraphe est formulé selon la position de *ka* par rapport à d'autres constituants avec lesquels il partage la même concaténation. Chaque libellé contient un indicateur de désambiguïsation pressenti du *ka* de la possession qui reste à vérifier. Juste après le titre du paragraphe, il est formulé l'hypothèse spécifique respective. Au sujet de l'hypothèse, des termes de recherche sont fournis. Ces termes sont une sous-section qui suit celle de l'hypothèse dans le plan interne de chaque paragraphe, à commencer par le sous-titre 4. Les termes de recherche servent de source de vérification pour les hypothèses respectives. Ils sont très souvent des « unités corrélatives » en rapport avec *ka*. Une fois les termes de recherche communiqués au logiciel (voir § 1), les phrases contenant les occurrences des termes de recherche correspondants sont inventoriées. Cet inventaire est la liste à analyser dans la section « observation ». Au cours de l'observation de la liste, le comportement de *ka* est scruté de concert avec les autres termes de recherche.

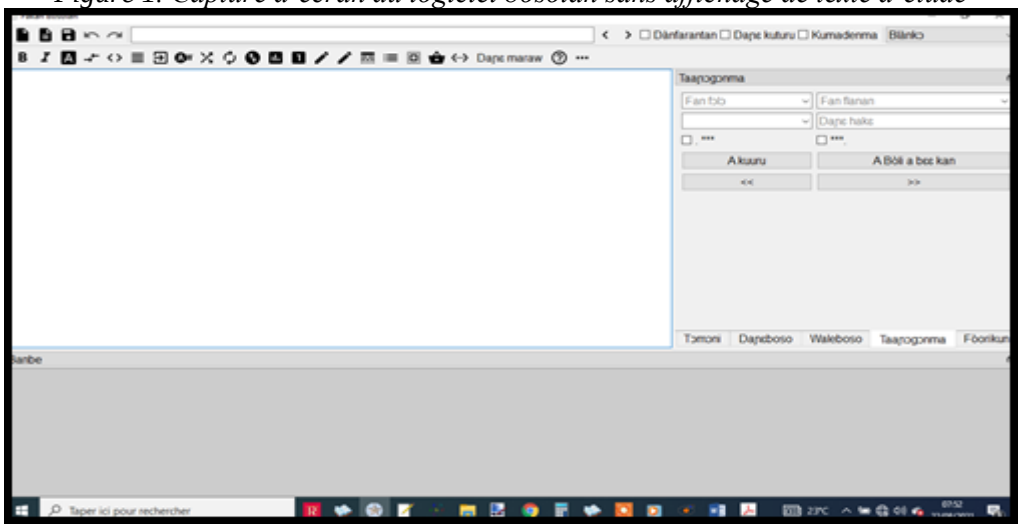
À la suite de la sous-section « observation », les résultats des observations sont fournis comme interprétation dans les sous-sections qui s'appellent « remarque ». Le nombre de remarques est cependant relatif à la complexité des détails de l'analyse. À la fin des sous-sections de chaque paragraphe, on retrouvera une sous-section « règle » qui suit les remarques. Cette dernière sous-section rapporte ce qu'on peut retenir comme enseignement de l'analyse faite. Cet enseignement est relatif à la reformulation de l'hypothèse émise dès le début en prenant en compte les résultats des observations faites. La règle explique comment les remarques peuvent être utiles à l'implémentation de l'hypothèse dans un programme informatique (Camara 1996 :12).

Toutefois, avant toute chose, il serait utile d'introduire l'outil de traitement Bosolan qui a servi dans notre travail.

1. Outil de traitement de l'environnement lexical : le logiciel Bosolan

Le programme *Bosolan* ayant servi de support au présent travail est un logiciel de traitement lexicographique que nous avons spécifiquement mis au point (Ballo, 2016) pour le bamanankan. Il est tout aussi adaptable aux langues africaines en plus de l'anglais et le français. Le logiciel est hors ligne. Il doit être installé sur des ordinateurs contenant Windows. Son interface d'utilisation est en bamanankan. Le menu y est assez fourni. Il comporte plus d'une trentaine d'options. Sa fonction principale est le traitement lexicographique simple et complexe. De ce fait, *Bosolan* est muni de plusieurs rubriques. Parmi les plus remarquables, il y a *Tɔmɔni* (extraction simple ou composite des occurrences d'un corpus), *Daɲɛboso* (rédaction de dictionnaire), *Waleboso* (traitement des verbes d'un corpus), *Sigannadaɲɛ* (filtre pour l'édition semi-automatique), *Taaɲɔgɔnma* (traitement des unités corrélatives du corpus).

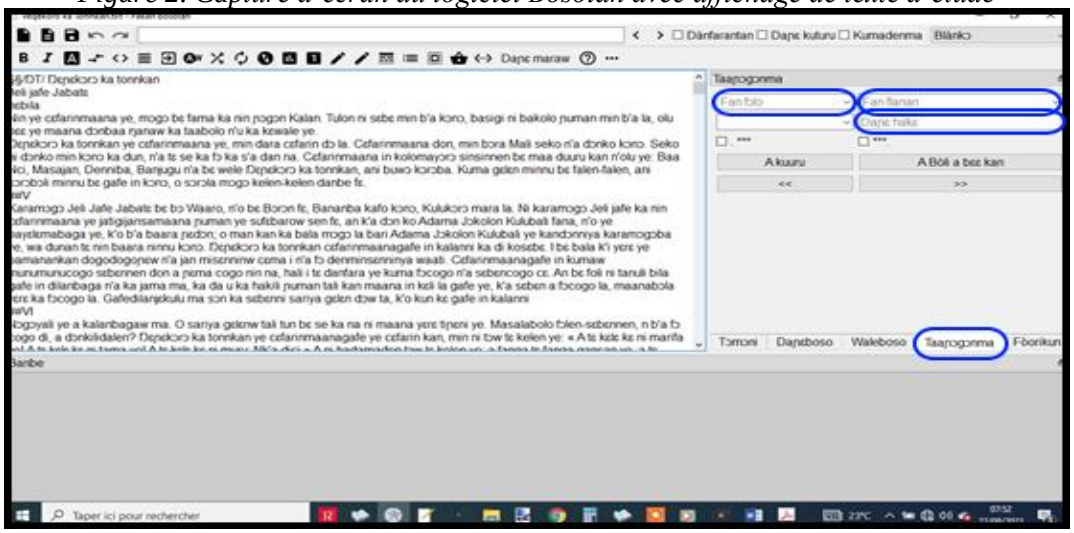
Figure 1. Capture d'écran du logiciel bosolan sans affichage de texte d'étude



L'option *Taaɲɔgɔnma* nous intéresse particulièrement dans ce travail, vu qu'elle a servi à faire, dans le cadre de notre corpus, le dépouillement des occurrences de l'environnement lexical (voir § 3) dans lequel le *ka* se trouve. Cet environnement lexical est repéré à l'aide des termes de recherche qu'on communique au logiciel. L'option *Taaɲɔgɔnma* (Corrélation) comporte, par défaut, des champs vides dans lesquels on saisit les termes de recherche qui sont sous la main : *fan fɔlɔ* (premier terme) + *furancɛlata* (intercalaire) + *fan flanan* (second terme) *daɲɛ hakɛ* (nombre de mots intercalés). Chaque champ, excepté celui de *Daɲɛ hakɛ*, est doté systématiquement d'une liste déroulante contenant les noms des catégories de mots pouvant se trouver dans un corpus. Lorsqu'il s'agit de repérer et d'extraire toutes les occurrences de *ka* suivies immédiatement par un nom potentiel dans un corpus, on voit alors la formule suivante dans les champs de *Taaɲɔgɔnma*: premier terme (*ka*) + intercalaire (0 occurrence) + second terme (nom potentiel). Les champs étant remplis par les termes de recherche cités, le lancement d'une telle commande fait le repérage et au besoin l'extraction des occurrences de *ka* dans un environnement occupé à droite de *ka* par un mot potentiellement de la catégorie des noms. Puisqu'on parle de zéro intercalaire entre le *ka* et le nom potentiel, il s'agit bien d'un *ka* immédiatement suivi par un nom potentiel. Le chiffre faisant office du nombre de mots intercalés varie selon le nombre de mots intercalés entre les deux bouts de la corrélation. Les champs ayant été remplis, les boutons de navigation (flèches) permettent de naviguer en avant ou en arrière. Ce déplacement indique, un à un, les emplacements des occurrences recherchées. A défaut, on peut cliquer sur le bouton qui lance l'extraction de toutes les occurrences. La liste des phrases contenant les termes assortis du nombre d'occurrences extraites s'affiche alors sur la console (banbe). Cette liste subit ainsi des observations dans la sous-section qui porte ce nom.

Dans cet article, il est question du vocable *ka*, surtout le *ka* de la possession. Quelles sont donc les données disponibles au sujet du *ka* ? Le paragraphe suivant apporte des réponses.

Figure 2. Capture d'écran du logiciel Bosolan avec affichage de texte d'étude



Les options ayant servi dans l'extraction des occurrences recherchées dans le présent travail sont en encadrées en couleur bleue

2. L'unité *ka* et le recoupement d'orthographe

Le *ka* dont il est ici question est celui de la possession. Il est appelé tigiyalan *ka* (Dnafla 1997, Diallo 2006, Dukure 2009) dans la métalangue bamanan. Le *ka* de la possession est l'homographe de trois autres *ka* de statut lexical distinct. On les rencontre tous en abondance dans un corpus bamanan : *ka* de l'infinitif, *ka* de l'impératif, *ka* qualificatif. Le couple de phrases suivant illustre bien l'occurrence des *ka*, y compris le *ka* de la possession, dans les énoncés bamanan :

U bɔra so kɔnɔ n'u ka (1) npalanw ye ka (2) kungo sira minɛ (Ils sont sortis de la maison étant munis de leur gibécière et se sont dirigés vers la forêt) ;
Zan min ka (3) jan ni Nci ye, o ko an ka (4) an teliya ka taa (Zan qui est plus grand que Ntji annonça qu'ils se dépêchent pour y aller).

Le *ka* (1) est celui de la possession pendant que (2) est celui de l'infinitif. Le (3) exprime le qualificatif tandis que (4) tient pour l'impératif. Or, la différenciation de ces fonctions pose problème au niveau du traitement du discours en bamanankan par l'intelligence artificielle. Cela passe par l'analyse de l'environnement lexical dans lequel chaque *ka* baigne. Cette analyse exploite l'entourage lexico-syntaxique de *ka* en cherchant à désambiguïser *ka* de toutes ses multiples catégories potentielles excepté la catégorie réelle annoncée par les constituants immédiats qui l'entourent. Ainsi, chaque occurrence de *ka* est confrontée à l'hypothèse émise. A travers un tri, les occurrences confirmant l'hypothèse reçoivent manuellement des marques qui les différencient des éventuelles occurrences qui l'infirment.

Au sujet « d'environnement lexical » abondamment utilisé dans notre travail, nous apportons des précisions dans la section qui suit.

3. Quelques références sur l'environnement lexical

L'environnement lexical est une notion exploitée dans le travail pour faire allusion à la combinatoire (Polguère 2008 : 40) des unités lexicales dans un énoncé. La littérature traite la question sous différentes terminologies et différents aspects. Le même Polguère (2008 : 106) emploie la notion de « lien lexical ». Lehmann (2008 : 85-86, 98) et Gaudin (2000 : 132, 160) utilisent plutôt celle de « environnement linguistique » pendant que Lérat (2010 : 69, 86, 122) parle d'« environnement syntaxique ou syntagmatique ». Dubois (2007 : 114) parle de « constituant immédiat » tandis que Jean-Michel Adam (2008 : 103) use de la notion de « liages du signifiant ».

Dans une phrase, les unités lexicales se suivent selon les lois de relations syntagmatiques (Lehmann 2008 : 23). Dans cette logique, chaque unité lexicale est toujours bornée des deux côtés par une autre unité dans la chaine de combinaison : *il cède le pouvoir à son fils*. Dans cette phrase, l'unité « pouvoir » est bornée à gauche par l'unité « le » tandis qu'elle est bornée à droite par l'unité « à ». Dans une telle position, l'unité « pouvoir », qui possède habituellement une catégorie flottant entre le nom et le verbe, se désambiguïse rien que par la présence du déterminant de nom « le » dans son environnement immédiat gauche. Dans un autre exemple (*Nous allons pouvoir procéder de la manière suivante*), bien qu'apparaisse la même unité lexicale « pouvoir », cette fois-ci cette dernière n'a pas la fonction de nom vu que son environnement immédiat a changé d'occupants. L'unité « pouvoir » y est bornée à gauche par l'unité « allons » et à droite par l'unité « procéder ». La présence de « allons » (verbe) dans la proximité immédiate gauche de « pouvoir » a fait de cet item un mot de la catégorie verbe. Sur la base de ces réflexions, nous disons que dans une suite de constituants immédiats, un constituant quelconque peut conférer un statut exclusif à un proche de catégorie habituellement ambigüe dans la langue. Cette idée reprend ce que postule Condamines (2005 : 44) dans son étude des *marqueurs de relations conceptuelles* : « Il s'agit d'éléments linguistiques, le plus souvent lexicaux ou lexico-syntaxiques, qui permettraient de repérer systématiquement une ou l'autre relation conceptuelle ». La même auteure oriente aussi la question de l'environnement lexical vers le repérage des unités terminologiques, surtout les marqueurs d'hyperonyme (Condamines 2007 : 48). Dans tous les cas, il s'agit de tirer profit de la position logique de l'unité sur la base du rapport qu'elle entretient avec les unités qui l'environnent dans la chaine. Un rapport qui confère le plus souvent un statut quelconque désambiguïsé et recherché pour ladite unité.

Les exemples abondent en bamanankan : dans l'énoncé « *Ŋolo fa ka moto binna* (la moto du père de Ngolo est tombée) », l'intelligence humaine classe facilement et à raison le *ka* parmi ceux de la possession. Ce classement correct du *ka* est fonction de la maîtrise du statut des unités (*fa* et *moto*) qui le bornent dans son environnement immédiat. Cependant, pour que cette maîtrise soit transférée de l'intelligence humaine à l'intelligence artificielle, il faudra mettre au point plusieurs procédés pouvant susciter un aiguillage souhaité chez la machine. Les paragraphes qui suivent traitent les différents environnements lexicaux qui ont permis la désambiguïsation des *ka* de la possession dans notre corpus.

4. Ka dans l'intercalaire de « y'a » et un nom

Hypothèse : tout *ka* faisant occurrence dans l'intercalaire de « y'a » et un nom s'exempte de tout autre statut excepté le statut du *ka* de la possession.

Termes de recherche : y'a + *ka* + nom

Observations[1]

ƉT ɲ.21 a y'a ka marifa ta ; (il a pris son fusil ;)
ƉT ɲ.130 Kabini Ala y'a ka so da, (depuis que son cheval est au monde,)
ƉT ɲ.169 Ni bilakoro y'a ka bilakorokuma fɔ, (si le gamin s'exprime en gamin,)
ƉT ɲ.175 U y'a ka bonya barikada. (ils ont loué sa gratitude.)
ƉT ɲ.246 o y'a ka ko bannen ye. (son cas est ainsi réglé.)

Remarque : le dépouillement des termes de recherche ***y'a + ka + nom*** a fourni 8 occurrences, soit 0,01% des 50 000 occurrences totales du corpus, 0,27% des 2 945 *ka* du corpus ou encore 1,03% des 774 *ka* de la possession. Exceptionnellement, l'ensemble des 8 occurrences confirme l'hypothèse, soit un pourcentage de 100%. Du moment où l'hypothèse a été confirmée partout dans le corpus, elle peut être érigée en règle d'où la formulation suivante.

Règle : tout *ka* faisant occurrence dans l'intercalaire de « y'a » et un nom se résigne exclusivement au statut du *ka* de la possession.

5. Ka dans l'intercalaire du pronom et du nom absolu[2]

Hypothèse : tout *ka* faisant occurrence dans l'intercalaire du pronom et du nom absolu se désambigüise de toutes ses autres catégories potentielles pour ne se résigner qu'à la seule catégorie du *ka* de la possession.

Termes de recherche : pronom + *ka* + nom absolu

Observations

ƉT ɲ.13 Ne kɔni ka sɔgɔmafɛfɔli ni n ka tilesekuncɛfoli ni n ka wulafɛfoli, (moi mon bonjour au réveil et mon bonjour à midi et mon bonsoir de l'après-midi,)
ƉT ɲ.27 A y'a ka marifa ta, (il a pris son fusil,)
ƉT ɲ.32 n bɛ taa n ka denbaya nɔ fɛ ka na, (je vais chercher ma famille pour l'amener,)

[1] La partie « observations » du traitement forme la liste des occurrences extraites du corpus à l'aide des termes de recherche. En revanche, vu que les listes sont longues d'une observation à une autre, un à deux échantillons-types sont retenus pour représenter les occurrences confirmant et infirmant l'hypothèse après l'analyse des occurrences inventoriées. Le nombre réel des occurrences inventoriées est quand même fourni dans la remarque (1) de chaque section. L'échantillon des occurrences infirmant l'hypothèse est marqué d'un astérisque.
[2] Voir le § 9 pour ce qui est de la définition de la notion de « nom absolu ».

ƉT ɲ.174 Ala m'a tora n ka taafe bila, (que Dieu me rende folle,)*
ƉT ɲ.256 Ne ka kirike da sokɛ kan, (lorsque je selle le coursier,)*

Remarque 1 : le dépouillement du corpus a inventorié 198 occurrences à propos des termes de recherche ***pronom + ka + nom absolu***, soit 0,39% des 50 000 occurrences totales du corpus, 6,72% par rapport aux 2 945 *ka* du corpus et 25,58% par rapport aux 774 occurrences de *ka* de la possession. Parmi les 198 occurrences, il en existe 188 qui confirment l'hypothèse, soit 0,37% des occurrences totales du corpus, 6,38% des 2 945 *ka* du corpus, 24,28% des 774 *ka* de la possession et enfin 94,94% des 198 occurrences inventoriées. Les détails relatifs aux infirmations sont fournis en remarque 2.

Remarque 2 : en regardant de près les 12 occurrences qui infirment l'hypothèse, le constat est vite fait : il s'agit des *ka* de l'impératif. Le facteur qui a occasionné les infirmations est la transitivité des verbes qui suivent le nom absolu. En effet, le verbe qui succède au nom absolu, directement ou non, est employé avec un antécomplément[1] qui est le nom absolu lui-même. Telle est l'explication de l'occurrence des *ka* de l'impératif dans la suite ***pronom + ka + nom absolu***. Donc, dans un tel environnement quadripartite (***pronom + ka + nom absolu + verbe à antécomplément***), il est difficile de trouver un *ka* de la possession bien qu'il soit précédé d'un pronom et suivi par un nom absolu. Afin que le *ka* soit un *ka* de la possession dans un tel environnement, il faut que la suite quadripartite soit précédée d'un verbe : ***verbe*** + ***pronom + ka + nom absolu + verbe à antécomplément***. Ainsi s'explique le fait que les 188 occurrences parmi les 198 inventoriées aient confirmé l'hypothèse.

Toutefois, l'hypothèse peut être établie en règle d'implémentation dans un programme informatique qui exécute le traitement artificiel du bamanankan d'où la formulation suivante.

Règle : tout *ka* faisant occurrence dans l'intercalaire du pronom et du nom absolu se désambigüise de toutes ses autres catégories potentielles pour ne se résigner qu'à la seule catégorie du *ka* de la possession. Cependant, vu que la règle comporte des exceptions, surtout au niveau des noms absolus qui font figure d'antécomplément d'un verbe de l'énoncé, son traitement systématique a besoin d'autres formulations plus précises pouvant écarter toute équivoque. Voici, ainsi, quelques règles plus exclusives découlant de la première :

- Tout *ka* faisant occurrence dans l'intercalaire de la suite tripartite pronom + ka + nom absolu se trouvant dans la proximité supérieure[2] d'un verbe s'exempte de tout autre

[1] Le verbe à antécomplément est un verbe admettant un complément qui se place avant lui « *ka* a fɔ : dire ». La métalangue bamanan l'appelle wale ɲɛdafama (Dnafla 1997 : 29). Le phénomène de verbe admettant un complément se plaçant avant lui-même n'est pas pertinent en français d'où le manque de terminologie prenant en compte la ramification dans la métalangue du français. Il n'est remarqué en français que dans les infimes cas de l'emploi de « en » comme pronom remplaçant son antécédent : « cette belle cérémonie, on en parle partout ». Le contraire du verbe à antécomplément est le verbe à postcomplément appelé wale kɔdafama dans la métalangue bamanan. Ce dernier est un verbe admettant un complément qui suit le verbe lui-même « *ka* jɔ a la : avouer ». Les phénomènes de verbe à antécomplément et de verbe à postcomplément sont très pertinents dans l'analyse du verbe bamanan.

[2] Le terme « proximité supérieure », avec comme contraire la « proximité inférieure », est utilisé dans le travail pour indiquer la position d'un mot à droite d'un autre mot. Autrement dit, le mot qui succède à un autre. Dans « Madu *ka* muru (le couteau de Madu) », « muru » est dans la proximité supérieure de « *ka* » tandis que « Madu » se trouve dans sa proximité inférieure.

statut pour se résigner au statut de ka de la possession : n'o tɛ cɛnin bɛ t'i ka bulonba wolonwula tiɲɛ dɛ (sinon, le petit va détruire tes sept vestibules) ;

- Tout *ka* faisant occurrence dans l'intercalaire du pronom et du nom absolu suivi par une ponctuation se désambigüise de toutes ses autres catégories potentielles pour ne se résigner qu'à la seule catégorie du ka de la possession : ne kɔni sera n ka so (je suis quand même arrivé chez moi) ;
- Tout *ka* faisant occurrence dans l'intercalaire du pronom et du nom absolu suivi par une postposition se désambigüise de toutes ses autres catégories potentielles pour ne se résigner qu'à la seule catégorie du ka de la possession : ne bɛ n ka ɲɔbufiyɛ la (je suis en train de vanner mon mil);
- Tout *ka* faisant occurrence dans l'intercalaire du pronom suivant un auxiliaire et du nom absolu se désambigüise de toutes ses autres catégories potentielles pour ne se résigner qu'à la seule catégorie du ka de la possession : a y'a ka marifa ta (il a pris son fusil);

Ces règles spécifiques paraissent très restrictives mais elles sont quand même très exclusives d'autant plus que les occurrences qui les appliquent parmi les 198 occurrences sont toutes totalement désambiguïsées en matière de statut en faveur du *ka* de la possession.

6. Le *ka* de la possession et l'élision

Hypothèse : toute occurrence de *ka* pouvant subir l'élision est d'office écartée du statut du *ka* possessif.

Termes de recherche : à la recherche de chaque occurrence du « k' » dans le corpus

Remarque : la vérification de l'hypothèse dans le corpus entier a identifié 945 *ka* avec apostrophe (k'), soit 1,89% des 50 000 occurrences totales du corpus, 32,08% des 2 945 « *ka*/k'» du corpus, 122,09% par rapport aux 774 *ka* de la possession. Aucune de ces occurrences de « k' » ne s'est révélée être un *ka* de la possession ayant pris la marque de l'élision. Cependant, dans le lexique général de la langue bamanan, surtout à l'oral, on entend souvent « *cɛ min y'a bugɔ o don, a* ***k'o*** *cɛ nalen filɛ* » pour dire sans élision « *cɛ min y'a bugɔ o don, a* ***ka o*** *cɛ nalen filɛ* (le type qui l'a frappé l'autre jour, voici ce dernier arrivé) ». Dans un tel environnement, le « k'o » dans l'énoncé ayant appliqué l'élision est bien le *ka* de la possession apostrophé. L'hypothèse est traduisible en règle d'où la formulation suivante.

Règle : toute occurrence de *ka* pouvant subir l'élision est d'office écartée du statut du *ka* possessif. Par contre, il faudra compter sur les éventuelles exceptions tel le seul cas rare mentionné dans la remarque 1.

7. *Ka* dans l'intercalaire de deux noms

Hypothèse : tout *ka* faisant occurrence entre deux mots de statut substantif est totalement désambiguïsé de ses autres statuts potentiels et se résigne au seul statut de marqueur de possession.

Termes de recherche : nom + *ka* + nom

Observations

ƉT ɲ.13 O ye kɛlɛmasadenw ka laada kɔrɔ ye? (cela est-il une vielle coutume chez les héritiers de guerriers ?)
ƉT ɲ.90 faama kun ma don Denniba ka so kɔnɔ. (sa majesté n'a mis pied dans la chambre de Denniba.)
ƉT ɲ.165 Masajan ka npogo donni ma diya maa o maa ye, (tous ceux qui n'ont pas été d'accord avec le tabassage de Masajan,)
ƉT ɲ.11 Denkɛ cɛ fila in nana ka na kodiy'a ye ka tɛmɛ denkɛ kɔrɔba in kan, (il a fini par avoir plus d'estime pour ces deux enfants que le fils ainé,)
ƉT ɲ.16 A sinna ka tɛmɛ ka bɔ a dɔgɔninw nɔ fɛ. (il sortit tout droit à la suite de ses jeunes frères.)

Remarque 1 : le dépouillement a fourni 185 occurrences des termes de recherche ***nom + ka + nom*** provenant du corpus, soit 0,37% des 50 000 occurrences du corpus, 6,28% des 2 945 *ka* du corpus, 23,90% par rapport aux 774 *ka* de la possession. Sur les 185 occurrences, il y figure 46 occurrences qui confirment l'hypothèse, soit 0,09% des 50 000 occurrences totales du corpus, 1,56% des 2 945 *ka* du corpus, 5,94% des 774 *ka* de la possession ou encore 24,86% des 185 occurrences inventoriées. En revanche, les facteurs qui invalident le reste 139 occurrences sont énumérés dans la remarque 2.

Remarque 2 : les facteurs qui ont joué en faveur de l'invalidation des 139 occurrences sont d'ordre orthographique. Les possibilités de recoupement qui existent entre les formes des unités de catégorie distinctes ont fait que plusieurs occurrences ont été inventoriées à tort sans qu'elles ne soient réellement de la catégorie nom. En substance, le *ka* existe dans l'intercalaire de plusieurs noms dont la morphologie recoupe d'autres mots de catégories distinctes. C'est le cas des homographes. Parmi ces *ka* invalidés figurent des *ka* impératifs, des *ka* infinitifs et des *ka* qualificatifs. Voici des exemples de *ka* entre deux mots de statut différent du nom mais qui sont des homographes de noms :

- *Dɔ wɛrɛ ka na : ka impératif entre adverbe et verbe homographes de noms (wɛrɛ et na);*
- *ka jugunin cɛkɔrɔba in sama ka bɔ : ka infinitif entre verbe et verbe homographes de noms (sama et bɔ);*
- *E tun ka kan ka sa : ka infinitif entre adjectif et verbe homographes de noms (kan et sa) ;*
- *mɔgɔninfin fɛrɛn ka di marisedɛsi fɛrɛn ye : ka qualificatif entre nom et qualificatif homographe de noms (fɛrɛn et di).*

L'hypothèse émise peut être érigée en règle d'implémentation dans un programme de traitement automatique du bamanankan.

Règle : le *ka* faisant occurrence entre deux mots de statut substantif se résigne en général à son seul statut de marqueur de possession. Les exceptions à cette règle sont entre autres :

- Le *ka* dans l'intercalaire de deux noms où il fait office de marqueur de verbe à antécomplément à l'impératif ;
- Le *ka* dans l'intercalaire de deux verbes si les deux verbes sont des homographes d'autres mots de la catégorie substantive et en plus si le verbe que *ka* introduit n'admet pas de complément qui se place avant le verbe lui-même.

8. Ka dans l'intercalaire d'un pronom et d'un nom

Hypothèse : tout *ka* faisant occurrence entre un pronom et un nom est désambigüisé de ses catégories potentielles excepté la catégorie de la possession.

Termes de recherche : pronom + *ka* + nom

Observations

ƉT ɲ.18 I bɛ ne ka saya ni n ka balo dɔn cogo di? (comment tu arrives à deviner ma mort et ma survie ?)
ƉT ɲ.137 Woro ni kɔgɔ de bɛ n ka ganfaa kɔnɔ. (il y a de la cola et du sel dans ma musette.)
ƉT ɲ.73 aw ka to bulon kɔnɔ yan. (vous restez dans le vestibule ici.)
ƉT ɲ.37 Nka sanni n ka fasolataa in labɛn, (mais avant que je ne prépare mon voyage au village,)
ƉT ɲ.76 i ka kan ka kɛ kulusi ye k'i cɛ sutura; (tu dois cacher les tares de ton conjoint comme un pantalon servant à camoufler les parties intimes;).

Remarque 1 : l'inventaire des termes de recherche ***pronom + ka + nom*** a dévoilé 412 occurrences du genre dans le corpus, soit 0,82% des 50 000 occurrences totales du corpus, 13,98% des 2 945 *ka* total du corpus et 53,22% par rapport aux 774 *ka* de la possession. L'analyse des 412 occurrences a fourni 308 occurrences qui confirment l'hypothèse émise, soit 0,61% des occurrences totales du corpus, 10,45% des 2 945 *ka* du corpus, 39,79% par rapport aux 774 *ka* de la possession ou encore 74,75% des 412 occurrences inventoriées. Cependant, parmi les 308 occurrences, il en existe quelques-unes dans lesquelles la particule « nin » (démonstratif) s'insère entre le *ka* et le nom qui le suit : *e ka* ***nin*** *su saba kɛra anw bolo kalo saba ye.*
Par contre, les occurrences qui infirment l'hypothèse ont des comportements propres qui sont précisés dans la remarque 2.

Remarque 2 : en regardant de près les 104 occurrences qui infirment l'hypothèse, il ressort que les *ka* à l'étude dans ces occurrences sont plutôt des *ka* de l'impératif (*/&) et du qualificatif (@) au détriment du *ka* de la possession. Les *ka* de l'impératif s'élèvent au nombre de 98 occurrences tandis que le *ka* qualificatif produit 6 occurrences.
À leur tour, les *ka* de l'impératif sont aussi catégorisés en fonction de leur transitivité. À ce propos, il a été identifié des *ka* impératifs s'accordant avec les verbes intransitifs (*) et des *ka* impératifs qui s'accordent avec des verbes à antécomplément (&).
Qu'il s'agisse des *ka* impératifs ou qualificatifs, leur présence est inattendue dans l'inventaire dans la mesure où il a été ordonné à la machine de n'extraire que les ***pronoms + ka*** suivi ***par un nom***. Après tout, il y eu la présence de *ka* suivi par ces fameux verbes et adjectifs. Cette présence incongrue de verbes et d'adjectifs dans l'inventaire s'explique par le facteur de correspondance d'orthographe entre les formes desdits verbes et desdits adjectifs et des noms : kuma (n) / kuma (v), *ka*n (n) / *ka*n (adj).
Par ailleurs, les seuls cas d'infirmation qui alignent immédiatement le nom après le *ka* tout en restant toujours un *ka* impératif se trouvent être les occurrences dans lesquelles le *ka* impératif annonce un verbe à antécomplément. Dans de tels cas, l'antécomplément, un substantif, se place entre le *ka* impératif et son verbe comme si nous avions affaire à un *ka* de la possession :

e ka baramusoya sɔrɔ. L'hypothèse peut être établie en règle quelconque d'implémentation de programmes informatiques, d'où la formulation suivante.

Règle : tout *ka* faisant occurrence dans l'intercalaire d'un pronom et d'un nom est désambigüisé de toute catégorie potentielle exceptée la catégorie de la possession. Cependant, il reste indispensable de prêter une attention particulièrement aux occurrences infirmant l'hypothèse vu qu'elles sont des exceptions à la règle. Ces exceptions à la règle sont des cas de flottaison de catégorie énumérés dans la remarque 2. L'implémentation doit en tenir compte de sorte qu'elles ne nuisent pas à la qualité de l'automatisme que la règle servira à traiter.

9. *Ka* dans l'intercalaire de deux noms absolus

Nous employons le terme « nom absolu » pour faire allusion aux unités lexicales qui ne se présentent que dans la seule catégorie lexicale du nom dans le lexique général de la langue. En général, dans le lexique de la langue bamanan, il existe plusieurs unités admettant plusieurs catégories potentielles selon l'environnement syntaxique qui les emploie. Les exemples suivants sont fournis pour ne citer que trois catégories potentielles pour l'unité « kɔrɔ » :

- *kɔrɔ (vieillesse) nom « kɔrɔ ka na man di (la vieillesse est épuisante) ;*
- *kɔrɔ (agé / vieux / vielle) adjectif « a man kɔrɔ ni ne ye (il n'est pas plus âgé que moi).*
- *kɔrɔ (viellir, prendre de l'âge) verbe « a b'a la ka kɔrɔ ka t'a fɛ (il vieilli de plus en plus).*

En revanche, il existe des noms qui ne souffrent d'aucune ambiguïté au sujet de leur catégorie lexicale. Ces mots possèdent le statut exclusif de nom où qu'ils soient employés dans le lexique de la langue. Le logiciel de dépouillement est doté d'un fichier système contenant une liste ouverte de ces noms exclusifs. La liste contient déjà plus de 2000 noms du genre tel mɔgɔ (être humain), so (maison / cheval), ji (eau, liquide), nɛgɛso (vélo), bulon (vestibule). Le phénomène de la « catégorie multiple » pour la même unité n'est pas aussi développé en français qu'en bamanankan. Les quelques rares cas en français sont : devenir / le devenir, savoir / le savoir, pouvoir / le pouvoir, devoir / le devoir…
L'hypothèse suivante est formulée pour vérifier si le *ka* intercalé entre deux noms absolus est toujours un *ka* de la possession.

Hypothèse : toute occurrence de *ka* se produisant dans l'intercalaire de deux noms absolus se libère de toute autre catégorie potentielle pour se résigner à la catégorie du *ka* de la possession.

Termes de recherche : nom absolu + *ka* + nom absolu

Observations

ƉT ɲ.V Ɖɛɲɛkɔrɔ ka tonnkan ye cɛfarinmaana ye, (Ngeniekoro ka Tonnkan est une épopée de guerrier,)
ƉT ɲ.13 O ye kɛlɛmasadenw ka laada kɔrɔ ye? (cela est-il une vielle coutume chez les héritiers de guerriers ?)
ƉT ɲ.158 n'o ye Masajan ka baramuso ye, (qui est l'épouse favorite de Masajan,)

ƉT ɲ.239 ni ne ko n tɛ yɛlɛn bilakoro ka cɛgana minɛnen kan na, (si je dis que je ne monte sur les épaules d'un célibataire capturé par un gamin,)
ƉT ɲ.59 n'i b'a fɛ muso ka denɲuman wolo i ka du kɔnɔ, (si tu t'attends à ce que la femme enfante un enfant prodige chez soi,)*

Remarque 1 : la vérification de l'hypothèse selon laquelle le *ka* dans les termes de recherche ***nom absolu + ka + nom absolu*** se résigne à la catégorie du *ka* de la possession a fourni 265 occurrences brutes après le dépouillement. Ce chiffre donne 0,53% des 50 000 occurrences totales du corpus, 8,99% des 2 945 *ka* du corpus et 34,23% par rapport aux 774 *ka* de la possession. Parmi les 265 occurrences, 264 ont confirmé l'hypothèse émise, soit 0,52% des 50 000 occurrences totales du corpus, 8,96% des 2 945 occurrences de *ka* du corpus, 34,10% par rapport aux 774 *ka* de la possession ou encore 99,62% des 265 occurrences inventoriées. Les détails sur les facteurs qui ont prévalu dans l'infirmation d'une occurrence sont fournis dans la remarque 2.

Remarque 2 : en analysant l'unique occurrence infirmant l'hypothèse, on trouve que le *ka* y est impératif au lieu d'y être infinitif en ce sens où la phrase exprime un souhait. Par contre, un facteur quelconque ne se précise pas vu que les infirmations ne sont pas nombreuses pour avoir plusieurs échantillons d'un même cas. Cependant, il est à noter qu'il y a la présence d'un verbe à antécomplément juste après le nom suivant *ka*. Alors, on peut affirmer qu'un risque potentiel d'infirmation peut être la présence d'un verbe à antécomplément potentiel dont l'antécomplément est un nom absolu.

Remarque 3 : parmi les occurrences confirmant l'hypothèse, 224 d'entre elles sont une répétition de *Ŋɛɲɛkɔrɔ ka Tonnkan* (noms du premier héros du roman qui a servi de corpus) et 11 autres sont celles de *Denniba ka Buwɔ* (noms du second héros du roman qui a servi de corpus). Un échantillon a donc été pris pour représenter chacune des répétitions. Il est envisageable d'ériger notre hypothèse en règle. Alors, la formulation suivante fait office de cette règle.

Règle : toute occurrence de *ka* se produisant dans l'intercalaire de deux noms absolus se libère de toute autre catégorie potentielle pour se résigner à la catégorie du *ka* de la possession. Cependant, la prudence doit être de mise lorsque nous sommes en présence d'un verbe à antécomplément potentiel juste après le nom absolu suivant le *ka*.

Conclusion

Ce traitement du *ka* de la possession du corpus Ŋɛɲɛkɔrɔ *ka* Tonnkan (Jafe 2007) a tout d'abord fait les décomptes de l'ensemble des 2 945 *ka/k'* du corpus afin d'isoler les seules occurrences relatives au *ka* de la possession (774) des occurrences sur les autres catégories de *ka* qui ne font pas l'objet d'analyse du présent travail. Ainsi, le travail ne portant que sur la seule catégorie du *ka* de la possession, les conclusions tirées sont bien relatives à cette dernière.

En effet, des indicateurs matériels ont été trouvés qui peuvent guider un programme informatique. Un programme informatique quelconque est dorénavant capable d'identifier automatiquement chaque *ka* de la possession où qu'il soit dans un texte. Cet acquis est donc

susceptible de s'appliquer à toute forme de traitements automatiques du bamanankan : assistance à la création terminologique, enseignement des langues assistés par ordinateur, aides au feuilletage (Camara 1996 : 812).

Pour ce faire, le travail a donc formulé 6 hypothèses sur le *ka* de la possession. Le *ka* de la possession ayant fait 774/2 945 (26,28%), il y a eu 5 hypothèses dans chacune desquelles plus de la moitié des occurrences a été confirmé. Parmi ces 5 hypothèses, exceptionnellement 2 hypothèses (hypothèses 1 et 3) ont été intégralement confirmées. Les occurrences de confirmation sont au nombre de 1 759 pour les 6 hypothèses de la catégorie.

Les résultats fournis par le présent travail viennent compléter quelques avancées numériques au sujet du bamanankan. Précisons que l'outil numérique est présent en bamanankan à travers des logiciels de traitement de texte, des sites web et des applications android. Il devient impératif d'appliquer les solutions de l'intelligence artificielle à chacune des ressources numériques disponibles sur le bamanankan. Surtout, les traitements de textes et les sites web ont besoin des astuces de l'intelligence artificielle afin d'imiter l'homme dans l'exécution de certaines tâches ordinairement exercées manuellement. Il est évident que l'exploitation de l'environnement lexical des occurrences du corpus bamanan est au moins une des clés de la résolution du problème de reconnaissance automatique des occurrences. C'est ce que nous espérons avoir mis en exergue dans ce travail.

Bibliographie

- Adam Jean-Michel (2008), La linguistique textuelle. Introduction à l'analyse textuelle des discours, Paris : Armand Colin.
- Camara E., Ndamba J., Nstadi C., Rey v., Véronis J. (1996), Traitement informatique des langues africaines : problèmes et perspectives. In : Moukeli P. (ed.) CARI'96 : actes du 3ème colloque africain sur la recherche en informatique = CARI'96 : proceedings of the 3rd African conference on research in computer science. Paris : ORSTOM, 810-819. (Colloques et Séminaires). CARI'96 : Colloque Africain sur la Recherche en Informatique = CARI'96 : African Conference on Research in Computer Science, 3, Libreville (GAB), 1996/10/09-16. ISBN 2-7099-1333-X.
- Condamines Anne (2007), L'interprétation en sémantique de corpus : le cas de la construction de terminologies, Revue Française de Linguistique Appliquée, N° 121, p. 39-52.
- Condamines Anne (2005), La linguistique de corpus et terminologie, in Langages n° 157, p. 36-47.
- Diallo Youssouf et al, (2006), Bamanankan mabɛn, Bamako, Donniya.
- Dnafla (1997), Bamanankan sariyasun, Bamako, Dnafla.
- Dubois Jean et al. (2007), Grand dictionnaire linguistique & sciences du langage, Paris, Larousse.
- Dukure Mamadu (2010), Sɛbɛnko bɛrɛ, Bamakɔ, Makdas sɛbɛnca, non édité.
- Gaudin, François et Guespin Louis (2000), Initiation à la lexicologie française : de la néologie aux dictionnaires, Bruxelles, Editions Duculot.
- Jafe Jeli (2007), Ŋɛɲɛkɔrɔ ka Tonnkan, Bamako, Edis.
- Lehmann Alise et Martin-Berthet, Françoise (2008), Introduction à la lexicologie : sémantique et morphologie, Paris, Armand Colin.
- Lerat Pierre (2010), Les langues spécialisées. Paris, PUF.
- Polguère Alain (2008), Lexicologie et sémantique lexicale : notions fondamentales, Montréal, PUM.