

RECONNAISSANCE AUTOMATIQUE DES 3 MA : LA METHODE DE L'ORDRE D'OCCURRENCE DES UNITES BAMANAN

Issiaka BALLO

Faculté des Lettres, des Langues, et des sciences du langage
Université des Lettres et des Sciences Humaines de Bamako

issiakaballo79@gmail.com

Abstract

This contribution concerns 3 Bamanan (bambara) lexical units which are homographs but of distinct lexical category. These are the "ma's" which are classified between the categories of auxiliary, noun and postposition. In human intelligence, there is no problem placing such and such an occurrence of "ma" in its respective category. On the contrary, in the artificial intelligence of the language, there is the huge problem of identifying each other in their respective class. Without a solution to this problem, it is impossible to order the machine to perform any automatic processing: machine translation (bamanan-français or bamanan-english), grammar checker. The method of the order of occurrence of lexical units in the corpus is therefore used in this work to identify material evidence serving as a point of differentiation between types of "ma". Therefore, these physical evidence can be machine ordered through their implementation in an automatic recognition program. The application of the units order of occurrence method is made possible by the Bosolan lexicographic analysis software.

Keywords : *artificial processing, automatic recognition, corpus, lexical category, order of occurrence*

Résumé

Cette contribution porte sur 3 unités lexicales bamanan (bambara) qui sont homographes mais de catégorie lexicale distincte. Il s'agit bien des ma qui sont classés entre les catégories d'auxiliaire, de nom et de postposition. Dans l'intelligence humaine, il ne se pose pas de problème pour situer telle ou telle des occurrences de ma dans sa catégorie respective. Par contre, dans l'intelligence artificielle de la langue, il se pose l'énorme problème d'identification des uns et des autres dans leur classe respective. Sans une solution à ce problème, il est impossible d'ordonner à la machine de procéder à un traitement automatique quelconque : traduction automatique (bamanan-français ou bamanan-english), correction syntaxique. La méthode de l'ordre d'occurrence des unités lexicales dans le corpus est donc mise à profit dans ce travail pour identifier des preuves matérielles servant de point de différenciation entre les ma. Par conséquent, ces preuves matérielles peuvent être ordonnées à la machine à travers leur implémentation dans un programme de reconnaissance automatique. L'application de la méthode de l'ordre d'occurrence des unités est rendue possible grâce au logiciel de dépouillement lexicographique Bosolan.

Mots-clés : *catégorie lexicale, corpus, ordre d'occurrence, reconnaissance automatique, traitement artificiel*

Introduction

Pour introduire ce travail sur la reconnaissance automatique des unités lexicales du bamanankan (bambara), il est possible de faire un bref rappel de ceux ou celles qui ont ouvert les voix de cette informatisation des langues africaines. La citation qui stipule que « *la réalisation d'outils informatiques pour les langues africaines semble être un enjeu important pour la place de ces langues dans le monde moderne basé sur les communications, et pour celles des cultures qu'elles représentent* » (Camara, 1996 :817) est assez évocatrice.

Vu l'importance de l'enjeu longtemps chanté par les chercheurs, il convient d'œuvrer dans ce sens pour faire des langues africaines des langues aussi dynamiques dans le numérique que leurs homologues qui ont pris le devant dans le processus de cette informatisation.

Le bamanankan du Mali est singulièrement une langue qui regorge des homographes. Dans un tel état, il est difficile d'ordonner automatiquement et ou systématiquement à la machine d'exécuter un traitement artificiel et avoir gain de cause. A titre illustratif, les occurrences de *ma* dans un corpus dont il est question dans ce travail sont difficilement distinguables selon les trois différentes catégories lexicales auxquelles ils répondent : auxiliaire, nom, postposition.

En tant qu'auxiliaire, *ma* est l'accessoire des verbes qui sont conjugués à l'accompli négatif. Le positif desdites conjugaisons se fait plutôt avec des morphèmes liés (-ra, -la, -na) ou libres (*ye*, *mana*) : *Madu ma boli* (négatif) donne *Madu bolila* (positif) : *Madu a fui*. Lorsqu'il est employé comme nom, il fait allusion à parent, seigneur, Dieu d'où il s'écrit souvent avec la majuscule à l'initiale : *An Ma ye Ala ye* : *Dieu est notre seigneur*. Enfin, son emploi postposition suggère qu'il est dans une relation tripartite *verbe + complément + postposition*. Dans ce trio, tout verbe à post-complément dont l'occurrence exige sa présence respective l'emploi à la suite du complément correspondant : *u y'a minε a tεgε ma* : *ils l'ont pris par les mains*. Par conséquent, il faut une analyse plus profonde pouvant mettre à la disposition du programmeur, des différenciateurs matériels qu'il implémente pour orienter le programme rien que vers la catégorie réellement employée dans telle ou telle position dans le corpus à traiter artificiellement.

Le travail procède au dépouillement d'un corpus à la recherche des énoncés contenant les occurrences de *ma* selon leur ordre d'apparition dans les énoncés respectifs. Ce dépouillement se fait à l'aide d'un logiciel (*Bosolan*) créée sur mesure pour le traitement lexicographique.

Chaque dépouillement est lancé à la suite de la formulation d'un sous-titre hypothétique et son hypothèse. L'hypothèse est aussi munie de termes de recherche pouvant être instruits à la machine à travers des options spécifiques du logiciel *Bosolan*. Le logiciel fourni donc un inventaire des occurrences relatives aux termes de recherche en guise de réponse à l'instruction donnée. Les occurrences de l'inventaire sont ensuite passées sous les loupes de l'analyse et de vérifications des *ma* qui confirment ou

infirmement l'hypothèse émise.

Les statistiques sont fournies au sujet du nombre des occurrences confirmant l'hypothèse dans une première remarque. D'autres remarques, s'il y a lieu, rapportent des détails sur les infirmations ainsi que sur la potentialité de déduction d'une règle issue de l'hypothèse. La règle est formulée tout en tenant compte des infirmations, s'il en existe, en tant qu'exceptions à la règle. A la fin de l'analyse, un commentaire met en relief comment la règle peut offrir comme résultat concret une fois implémentée dans une application informatique quelconque.

Les sous-titres hypothétiques explorés sont regroupés selon les 3 catégories de *ma* se trouvant dans le corpus. Dans la conclusion, les statistiques synthétiques sont données sur le nombre de *ma* ayant été exploré par la vérification des différentes hypothèses. C'est dire que les occurrences de *ma* dans le corpus qui n'auront pas la chance d'être couvertes dans les limites des hypothèses du présent travail ne figureront pas dans les statistiques fournies. Il faudra d'autres hypothèses pour couvrir ces *ma* non couverts.

1. L'auxiliaire *ma*

Le *ma* en tant que l'auxiliaire verbal s'appelle *waledɛmɛnan* (DNAFLA, 1997 :24) ou *walelan* (Dukure, 2009 :14) dans la métalangue du bamanankan. Cet auxiliaire s'emploie pour exprimer la négation chez les verbes bamanan conjugués à l'accompli : *ka taa* : affirmatif : *A taara* (il est parti) – Négatif : *A ma taa* (il n'est pas parti). Sa position est toujours antérieure à celle du verbe auquel il se rapporte. Du point de vue ton, l'auxiliaire *ma* possède le ton lexical haut. Il peut s'élider en présence d'une voyelle qui lui suit pour devenir « *m'* ». Les sous-titres hypothétiques suivants et leurs hypothèses respectives sont formulés pour distinguer sans doute les occurrences de l'auxiliaire *ma* des autres occurrences possibles de *ma* qui ne sont pas forcément des auxiliaires verbaux. Les traits distinctifs qui servent à démarquer les auxiliaires *ma* des autres *ma* pourront alors faire office de preuves matérielles pouvant être mises à profit dans l'implémentation de programmes de traitement artificiel du bamanankan.

1.1 *ma* en position de deuxième unité après la ponctuation

Hypothèse : tout *ma* en position de deuxième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*.

Termes de recherche : ponctuation + 1 intercalaire + *ma*

Observations : la liste des occurrences inventoriées du corpus pour observation n'est pas intégralement fournie. En raison des contraintes du volume exigé dans la contribution, un échantillon contenant 3 à 7 occurrences est fourni à titre illustratif. Ces occurrences contiennent des exemples de confirmation et d'infirmerie.

NT p.14 A ma kuma.

NYT p.24 Sogo ma ye,
NYT p.73 Ne ma jiri ta,

Remarque 1 : le dépouillement du corpus (50 000 occurrences absolues) à l'aide des termes de recherche à inventorié 163 occurrences contenant *ma* sur 757 occurrences absolues de *ma* dans le corpus. L'ensemble de ces 163 occurrences de *ma* confirment l'hypothèse émise. En voici quelques pourcentages :

163/163 (100%)

163/757 (21,53%)

163/50 000 (0,32%)

Remarque 2 : l'hypothèse émise peut être érigé en règle quelconque pouvant être implémenté dans un programme informatique au bénéfice du traitement automatique du bamanankan.

Règle : tout *ma* en position de deuxième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*.

Commentaire : la règle permet d'orienter l'implémentation des occurrences de *ma* d'un corpus exclusivement vers le sens du *ma* auxiliaire. Dans ce cas, lorsqu'on incorpore la règle à un logiciel de traduction automatique, le logiciel traduira systématiquement chaque *ma* en rapport avec la règle avec son équivalent dans l'autre langue : *n'avoir pas, n'être pas* (bamanankan-français) ou encore *have not, did not* (bamanankan-english). La règle est alors capable de traiter jusqu'à 21,53% les *ma* présents dans un corpus.

1.2. ma en position de troisième unité après la ponctuation

Hypothèse : tout *ma* en position de troisième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*.

Termes de recherche : ponctuation + 2 intercalaire + *ma*

Observations

NYT p.51 U si ma se ka basa faga.

NYT p.137 Nka n ma maa sɔrɔ ka taa o di jeliw ma.

*NYT p.16 A ko Ma,**

*NYT p.184 Buwɔ ko Ma,**

Remarque 1 : le dépouillement du corpus (50 000 occurrences absolues) à l'aide des termes de recherche à fait l'inventaire de 79 occurrences contenant *ma* sur 757 occurrences absolues de *ma* dans le corpus. Parmi ces 79 occurrences, 67 confirment l'hypothèse tandis que les 12 (*) occurrences restantes l'infirmement. De ce fait, il en est ressorti les pourcentages ci-après :

67/79 (84,81%)

67/757 (8,85%)

67/50 000 (0,13%)

Remarque 2 : les 12 occurrences ayant infirmé l'hypothèse contiennent des noms *ma*. Ces occurrences peuvent être isolées systématiquement des

occurrences de confirmation à l'aide des critères matériels qui sont :

- *Se sont des ma noms propres. Ils s'écrivent toujours avec la lettre initiale en majuscule ;*
- *Il s'agit des ma qui sont toujours suivis d'une virgule en respect aux normes rédactionnelles parce que le ma est la dernière unité d'un syntagme qui précède un autre syntagme qui lui est subordonné ;*
- *Se sont des ma qui sont toujours précédés de l'occurrence de « ko » (verbe dire).*

Remarque 3 : l'hypothèse est traduisible en règle d'implémentation dans un programme de traitement automatique du bamanankan d'où la formulation suivante.

Règle : tout *ma* en position de troisième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*. Par contre, il faut tenir compte des exceptions à la règle. Ces exceptions sont détaillées dans la remarque 3.

Commentaire : le programmeur peut ordonner à la machine d'adopter tel ou tel comportement envers les *ma* de la troisième position après une ponctuation. S'il s'agit d'une aide à la traduction, il va implémenter la règle dans le programme pour inciter la machine à traduire automatiquement *ma* en « *ne + auxiliaire respective + pas* » en français ou en « *did not, have not, ...* » en anglais. La machine aura exécuté cette tâche systématiquement en tenant compte des exceptions parallèlement programmées. En se référant aux pourcentages obtenus en remarque 1, la règle pourra ainsi traiter automatiquement le cas de 8,85% des 757 *ma* totals du corpus ou 0,13% des 50 000 occurrences absolues du corpus.

1.3 *ma* en position de quatrième unité après la ponctuation

Hypothèse : tout *ma* en position de quatrième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*.

Termes de recherche : ponctuation + 3 intercalaire + *ma*

Observations

NYT p.13 ne yafar'i ma,

NYT p.49 fɛn min ɲogɔn ma da.

*NYT p.16 n ko a ma,**

*NYT p.53 u k'o ma ko bolo naani.**

*NYT p.108 Sira dir'aw ma sa de.**

Remarque 1 : le dépouillement du corpus (50 000 occurrences absolues) à l'aide des termes de recherche à fait l'inventaire de 55 occurrences contenant *ma* sur 757 occurrences absolues de *ma* dans le corpus. Parmi ces 55 occurrences, 27 ont confirmé l'hypothèse tandis que les 28 (*) occurrences restantes l'ont infirmé, d'où les pourcentages ci-après :

27/55 (49,09%)

27/757 (3,56%)

27/50 000 (0,05%)

Remarque 2 : les 28 occurrences réfractaires à l'hypothèse contiennent

des postpositions *ma*. Ces occurrences peuvent être isolées systématiquement des occurrences de confirmation à l'aide des critères matériels qui sont :

- *La plupart de ces ma sont bornés par une ponctuation, critère qui n'est jamais rencontré chez l'auxiliaire ma, faisant qu'ils sont systématiquement couverts par une autre règle (voir sous-titre 3.1) ;*
- *Les quelques-uns qui ont une unité entre eux et la ponctuation suivante se rapportent à la règle du sous-titre 3.2 si et seulement si cette unité porte l'une des catégories lexicales suivantes : adverbe de lieu ou de temps, idéophones ;*
- *Les restes des occurrences réfractaires sont toutes prises en compte par la règle du sous-titre 1.4.*

Remarque 3 : l'hypothèse peut être érigée en règle concrète à implémenter dans un programme de traitement artificiel du bamanankan.

Règle : tout *ma* en position de quatrième unité après une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de l'auxiliaire *ma*. Par ailleurs, les quelques exceptions à la règle sont détaillées dans la remarque 2. La prise en compte de ces exceptions dans le programme éventuel est nécessaire pour orienter le traitement vers les seules occurrences dont l'appartenance à la catégorie de l'auxiliaire ne souffre pas de doute.

Commentaire : lorsqu'on incorpore la règle dans un programme quelconque de traitement automatique, elle peut résoudre le problème à hauteur de 3,56% de l'ensemble des *ma* relatifs à chaque corpus comportant 757 occurrences absolues de *ma*. Il se trouve que dans une telle position, le *ma* n'est jamais un *ma* qui s'écrit avec « M » majuscule comme le cas du nom *ma* qui exprime Dieu. Si par mégarde le *ma* de la présente règle avait été écrit dans un corpus avec un « M » majuscule, le correcteur syntaxique pourrait venir au secours en considérant l'occurrence à raison comme une faute. Parce que, le programme d'aide à la correction syntaxique saura qu'un *ma* (Dieu) avec « M » majuscule ne fait pas occurrence dans une telle position. L'activation de l'option correcteur va détecter la faute et par conséquent, le redacteur va restituer le « m » minuscule à sa place.

1.4ma précédant deux occurrences avant la ponctuation

Hypothèse : tout *ma* qui précède deux occurrences avant une ponctuation se désambiguïse de toute autre catégorie pour se résigner à la catégorie de l'auxiliaire *ma*.

Termes de recherche : *ma* + 2 intercalaires + ponctuation

Observations

NT p.204 A ma maa fa,

*NT p.9 u tun b'a f'olu ma ko jadɔw!**

*NT p.49 U sera wolojɛba in ma tuma min,**

*NT p.161 n'o kɛra olu ma saya ye,**

*NT p.252 Ncɛnkolonbala Makan y'a denmuso d'e ma furu la.**

Remarque 1 : le dépouillement du corpus (50 000 occurrences) à l'aide des termes de recherche à aboutit à l'inventaire de 77 occurrences contenant *ma* sur 757 occurrences absolues de *ma* dans le corpus. Parmi ces 77 occurrences, 62 (*) confirment l'hypothèse tandis que les 15 occurrences restantes l'infirmant. De ce fait, il en est ressorti les pourcentages ci-après :

62/77 (80,51%)

62/757 (8,19%)

62/50 000 (0,12%)

Remarque 2 : dans les 15 occurrences infirmant l'hypothèse, le *ma* est plutôt celui de la postposition et non celui de l'auxiliaire. En plus, le verbe de la partie de phrase qui emploie *ma* admet au moins deux syntagmes postpositionnels. Les constituants du premier de ces syntagmes se terminent à *ma* pour que les constituants du second commencent et qui se terminera aussi par une postposition respective. Les facteurs qui se sont réunis pour que le *ma* soit celui de la postposition dans un tel environnement sont entre autres :

- La présence de « *ko* » suivant directement *ma* dans la suite ;
- La présence d'un adverbe de temps qui suit directement *ma* et le tout suivi d'un mot de renforcement surtout un démonstratif (*min*) ou encore un adverbe de type « *sinsinnan* » (*bani, dεε...*) dans la suite ;
- La présence de deux adverbes de type « *sinsinnan* » qui se suivent comme intercalaire entre *ma* et la ponctuation ;
- La présence d'un adverbe de lieu et sa postposition respective qui suivent directement *ma* dans la suite ;

Remarque 3 : l'hypothèse peut être érigée en règle quelconque pouvant être implémentée dans un programme de traitement automatique du bamanankan.

Règle : tout *ma* qui précède deux occurrences avant une ponctuation se désambiguïse de toute autre catégorie pour se résigner à la catégorie de l'auxiliaire *ma*. Par contre, les facteurs provoquant l'occurrence d'un *ma* non auxiliaire dans une telle position sont fournis dans la remarque 2.

1.5ma élidé en « m' »

Hypothèse : tout *ma* ayant subi l'élimination est exclusivement un auxiliaire *ma*.

Termes de recherche : extraction de chaque proposition de phrase contenant l'occurrence de « m' »

Observations

NT p.14 *A ko ne m'a fo i kana n fo bilen wa?*

NT p.57 *u si m'a ta k'a tije.*

NT p.175 *an m'i jɔgɔn ye!*

Remarque 1 : le dépouillement du corpus (50 000 occurrences absolues) à l'aide des termes de recherche à aboutit à l'inventaire de 87 occurrences contenant le *ma* élidé sur 757 occurrences absolues de *ma*

dans le corpus. L'ensemble de ces 87 occurrences confirme l'hypothèse émise. De ce fait, il en est ressorti les pourcentages ci-après :

87/87 (100%)

87/757 (11,49%)

87/50 000 (0,17%)

Remarque 2 : l'hypothèse émise est donc traduisible en règle concrète pouvant être implémenté dans un programme de reconnaissance automatique du bamanankan.

Règle : tout *ma* ayant subi l'élision est exclusivement un auxiliaire *ma*.

Commentaire : une telle règle incorporée à un programme donne comme résultat la reconnaissance systématique de tout « m' » qui est réellement un *ma* élidé comme étant exclusivement un auxiliaire *ma*. Conséquemment, le programme le traitera à raison dans un correcteur syntaxique, en « m' » minuscule si par inadvertance il avait été écrit avec un « M' » majuscule. Cela faisant, l'automatisme pourra traiter 0,17% des 50 000 occurrences absolues d'un corpus comme le souligne les statistiques fournies en remarque 1.

2. Le nom *ma*

Dans la phrase « *A ma sogo dun* (Il n'a pas mangé de la viande) », le *ma* est un auxiliaire comme il a été évoqué dans la section précédente. Par contre, dans la phrase « *An ma ye Ala ye* (Dieu est notre seigneur) », le *ma* est un nom. Dans la présente section, des sous-titres hypothétiques assortis de leurs hypothèses respectives sont vérifiés pour identifier exclusivement les occurrences des *ma* de catégorie nom d'éventuelles occurrences appartenant à d'autres catégories lexicales. La métalangue bamanan appelle le nom (substantif) *togo* (Diallo, 2006 :66). Il possède un ton lexical haut.

2.1 *ma* avec « M » majuscule à l'initial du mot

Hypothèse : tout *ma* dont la lettre initiale « m » est écrite en majuscule « M » est exclusivement de la catégorie des noms.

Termes de recherche : recherche simple des *ma* en respectant la casse.

Observations

NT p.76 Ma, n y'a mɛn.

NT p.85 Hali ni ne ni don, munna ne t'o d'e Ma?

NT p.184 Buwo ko Ma, n y'i ka kuma mɛn, n tɛ kuma.

Remarque 1 : le dépouillement du corpus de 50 000 occurrences absolues à l'aide des termes de recherche a fait l'inventaire de 30 occurrences relatives à l'hypothèse sur 757 occurrences absolues de *ma* dans le corpus. L'ensemble des 30 occurrences ont confirmé l'hypothèse. Alors, les pourcentages suivants se dégagent.

30/30 (100%)

30/757 (3,96%)

30/50 000 (0,06%)

Remarque 2 : l'hypothèse qui stipule que tout *ma* qui commence avec la lettre majuscule est exclusivement un nom peut être érigée en règle d'implémentation dans un programme de traitement automatique du bamanankan.

Règle : tout *ma* dont la lettre initiale « m » est écrite en majuscule « M » est exclusivement de la catégorie des noms.

2.2ma en position de première unité après la ponctuation

Hypothèse : tout *ma* en position de première unité suivant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie du nom *ma*.

Termes de recherche : *ma* en tant que première unité après la ponctuation : 9 occurrences

Observations

NYT p.76 *Ma,*

NYT p.78 *Ma,*

NYT p.173 *Ma n y'a mɛn.*

Remarque 1 : le dépouillement du corpus de 50 000 occurrences absolues à l'aide des termes de recherche a fait l'inventaire de 9 occurrences relatives à l'hypothèse sur 757 occurrences absolues de *ma*. L'ensemble des 9 occurrences confirment l'hypothèse. De ce fait, les pourcentages suivants se sont dégagés.

9/9 (100%)

9/757 (1,18%)

9/50 000 (0,01%)

Remarque 3 : l'hypothèse est traduisible en règle concrète d'implémentation dans un programme informatique sur le traitement automatique du bamanankan.

Règle : tout *ma* en position de première unité en partant d'une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie du nom *ma*.

3. La postposition *ma*

La liste des postpositions (kɔbila / bilankɔ) bamanan contient un *ma* (Dukure, 2008). Ce *ma*, de ton lexical bas (Vydrine, 1999), est assez abondant dans le lexique de la langue bamanan. Dans la phrase « *Fanta ma ji di dunankɛ ma : Fanta n'a pas donner à boire à l'invité* », la première occurrence de *ma* en partant du sujet est un auxiliaire pendant que la seconde est une postposition. La postposition bamanan est une particule qui fait le rapport entre le verbe et le post-complément (*verbe admettant un complément qui suit le verbe lui-même* « *ka jɔ a la (reconnaitre)* ») auquel il s'accorde. C'est pour dire que la postposition ne fait occurrence dans un énoncé que lorsque ledit énoncé

contient un verbe à post-complément (Dukure, 2009 :24). La présente section analyse des hypothèses sur l'exclusivité des *ma* en faveur de la catégorie de la postposition.

3.1 *ma* en position de dernière unité avant la ponctuation

Hypothèse : tout *ma* en position de dernière unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*.

Termes de recherche : *ma* + 0 intercalaire + ponctuation

Observations

NYT p.23 A taara se fuga kenɛba do ma.

NYT p.34 i tɛ kolon kemɛ di n ma!

NYT p.186 a ko Jelicsɛkɔɔɔba ma:

*NYT p.16 Ma!**

*NYT p.184 Buwɔ ko Ma,**

Remarque 1 : dans le corpus d'environ 50 000 occurrences absolues, le dépouillement a fourni l'inventaire de 255 occurrences de *ma* dans la position cible sur 757 *ma* absolus du corpus. Sur ces 255 occurrences, 226 occurrences ont confirmé l'hypothèse tandis que les 29 (*) occurrences restantes l'ont infirmé. Par conséquent, il en ressort les pourcentages ci-après :

226/255 (88,62%)

226/757 (29,85%)

226/50 000 (0,04%)

Remarque 2 : les 29 occurrences en contradiction avec l'hypothèse émise contiennent toutes des noms *ma*. Ces *ma* peuvent être distingués des *ma* de confirmation de part la lettre majuscule à leur initiale.

Remarque 3 : l'hypothèse est traduisible en règle concrète au bénéfice du traitement artificiel du bamanankan d'où la formulation suivante.

Règle : tout *ma* en position de dernière unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*. Les exceptions à cette règle sont clarifiées dans la remarque 2.

Commentaire : suite à la mise en évidence de cette règle, un traducteur automatique peut enfin considérer à raison toute occurrence de *ma* de la règle comme une postposition. Par conséquent, le traducteur pour traduire systématiquement *ma* en son équivalent français ou anglais déjà incorporé dans un fichier système. La règle peut déjà traiter jusqu'à hauteur de 29,85% *ma* dans un corpus comportant 757 occurrences absolues de *ma* en se référant aux pourcentages en remarque 1.

3.2. *ma* en position de deuxième unité avant la ponctuation

Hypothèse : tout *ma* en position de deuxième unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*.

Termes de recherche : *ma* + 1 intercalaire + ponctuation

Observations

1. *NYT p.50 an k'a ma kulubali,**
2. *NYT p.19 A ko na marifa kɔrɔnin di n ma dɛ!**
3. *NYT p.40 tiɲɛ yɛɛ la i filila n ma wa?**
4. *NYT p.82 n'i y'i ka ko bɛɛ lajɛlen to cew ma bi,**
5. *NYT p.189 a taar'a yɛɛ di gadaw ma yen.**
6. *NYT p.44 n'a ma sa,*
7. *NYT p.24 a ma bɔ,*

Remarque 1 : Dans le corpus d'environ 50 000 occurrences absolues, le dépouillement a fait l'inventaire de 94 occurrences de *ma* dans la position cible. Sur les 94 occurrences inventoriées relatives à l'hypothèse, seulement 30 (*) occurrences ont confirmé l'hypothèse tandis que les 64 autres occurrences l'ont infirmé. Il est à noter que le nombre d'occurrences absolues de *ma* s'élève à 757. Il en ressort donc les pourcentages ci-après :

30/94 (31,91%)

30/757 (3,96%)

30/50 000 (0,06%)

Remarque 2 : l'occurrence « *NYT p.44 n'a ma sa,* » infirme bien l'hypothèse. Par contre, dans le lexique général de la langue, la même suite peut vouloir dire « *ka na mɔgɔ ma sa = ka n'a ma sa,* » avec un « *sa* » adverbe et non verbe. Dans une telle occurrence, la suite confirme l'hypothèse vu qu'il sera plutôt question de la postposition *ma* et non celle de l'auxiliaire *ma*. Ladite occurrence prête donc à confusion et par conséquent, mieux vaut l'écarter d'une prise en charge automatique et systématique. L'occurrence n'est pas un cas isolé. Donc, il doit être réservé un traitement non systématique à l'ensemble des cas de figure du genre.

Remarque 3 : les occurrences qui ont confirmé l'hypothèse l'ont fait parce qu'elles ont bénéficié du concours de certains facteurs favorisant la confirmation. Voici ces facteurs identifiés :

- Si l'intercalaire entre *ma* et la ponctuation est un nom de famille : *Kulubali, ... ;*
- Si l'intercalaire entre *ma* et la ponctuation est un « *sinsinnan* (adverbe) » : *kɛ, dɛ, dɛɛ, koyi, mɛnɛ, kɔni, de ;*
- Si l'intercalaire entre *ma* et la ponctuation exclusive (point d'interrogation) est le mot interrogatif « *wa* » ;
- Si l'intercalaire entre *ma* et la ponctuation est un adverbe de temps : *bi, fɔb, waati, sisan ;*
- Si l'intercalaire entre *ma* et la ponctuation est un adverbe de lieu :

- *yen* ;
- Si l'intercalaire entre *ma* et la ponctuation est un « *kabalan* (idéophone) » : *tuu, borobara, giriwu* ;

Remarque 4 : des occurrences comme « *kε*, de » prêtent à confusion vu qu'elles sont des homographes d'autres occurrences de catégorie différentes. Une liste à exhaustivité progressive au sujet desdites occurrences doit être fournie en fichier systèmes en guise de différenciation.

Remarque 5 : l'hypothèse peut être reformulée en tenant compte des occurrences réfractaires pour avoir une règle précise. En voici une reformulation ci-après.

Règle : tout *ma* dans la suite *ma* + 1 intercalaire + ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma* si et seulement si l'intercalaire figure parmi les mots détaillés dans la remarque 3.

Commentaire : le résultat fourni par l'activation d'une option propre à cette règle dans un programme de traduction est la traction systématique de chaque occurrence de *ma* de la règle en son équivalent français ou anglais déjà traité dans une interface dédiée à cette tâche. L'option pourra exécuter automatiquement la tâche jusqu'à atteindre la hauteur de 31,91% de 94 occurrences absolues de la postposition *ma*.

3.3. *ma* en position de troisième unité avant la ponctuation

Hypothèse : tout *ma* en position de troisième unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*.

Termes de recherche : *ma* + 2 intercalaire + ponctuation

Observations

*NYT p.38 U tora k'a f'a ma ko Masajan,**

*NYT p.40 A nana se bulon wolonwula in cεmancejiri ma waati min,**

*NYT p.188 A ko n'ale denmuso di i ma furu la,**

*NYT p.252 Ncεnkolonbala Makan y'a denmuso d'e ma furu la.**

NYT p.83 aw ma nin ye?

Remarque 1 : dans le corpus d'environ 50 000 occurrences absolues, le dépouillement a fait l'inventaire de 78 occurrences de *ma* dans la position cible sur 757 occurrences absolues de *ma* dans le corpus. Sur ces 78 occurrences inventoriées relatives à l'hypothèse, seulement 11 (*) occurrences ont confirmé l'hypothèse émise tandis que les 63 autres occurrences l'ont infirmé. Il en ressort donc les pourcentages ci-après :

15/78 (19,23%)

15/757 (1,98%)

15/50 000 (0,03%)

Remarque 2 : les 15 occurrences ont confirmé l'hypothèse grâce à l'environnement lexical dans lequel les *ma* respectifs se trouvent :

- *ma* + *ko* (*appeler, qualifier*) + *nom d'appartenance à une profession ou un clan*. Dans certains cas, le « *ko* » peut s'élider ».
- *ma* + *adverbe de temps* ;

- *ma* + deuxième syntagme postpositionnel qui exige « la » comme postposition.

Remarque 3 : l'hypothèse peut être traduite en règle concrète à implémenter dans un programme de traitement artificiel du bamanankan.

Règle : tout *ma* en position de troisième unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*. Cependant, il convient de bien prendre en compte les facteurs (remarque 2) qui concourent à la plénitude de la règle pour ne pas tomber dans les exceptions multiples qu'elle possède.

3.4.ma en position de quatrième unité avant la ponctuation

Hypothèse : tout *ma* en position de quatrième unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*.

Termes de recherche : *ma* + 3 intercalaire + ponctuation

Observations

NT p.54 n'i ma den galabuma wolo,

*NT p.53 u k'o ma ko bolo naani.**

*NT p.113 E k'a fo ne ma i ka telin!**

*NT p.252 Tonnkan ko saya ka fisa a ma n'o ye!**

*NT p.260 ka ba in bannen to dugu ma dukεε cεmance rε,**

Remarque 1 : dans le corpus d'environ 50 000 occurrences absolues, le dépouillement a fait l'inventaire de 59 occurrences de *ma* dans la position cible sur 757 occurrences absolues de *ma* dans le corpus. Sur ces 59 occurrences inventoriées relatives à l'hypothèse, seulement 52 occurrences ont confirmé l'hypothèse émise tandis que les 7 (*) occurrences restantes l'ont infirmé. Il en ressort donc les pourcentages ci-après :

52/59 (88,13%)

52/757 (6,86%)

52/50 000 (0,10%)

Remarque 2 : les 7 occurrences qui ont infirmé l'hypothèse contiennent des postpositions *ma*. Elles ont infirmé l'hypothèse à cause des environnements lexicaux spécifiques dans lequel les *ma* respectifs se trouvent qui sont :

- *ma* + *ko* (appeler, qualifier) + nom d'appartenance à une profession ou un clan. Dans certains cas, le « *ko* » peut s'élider ».
- *ma* + une suite adjectivale du genre *ma* + (*ko*) + sujet + *ka/man* adjectif ;
- *ma* + *ni* + *o* + *ye* précédé de la structure sujet + *ka fisa* + complément + *ma* ;
- *ma* + deuxième syntagme postpositionnel qui exige « *la, na,rε* » comme postposition.

Remarque 3 : l'hypothèse peut être reformulée en règle précise à implémenter dans un programme de traitement artificiel du bamanankan.

Règle : tout *ma* en position de quatrième unité avant une ponctuation se libère de toutes ses catégories potentielles pour se résigner à la catégorie de la postposition *ma*. En revanche, il faudra tenir compte des cas d'exceptions qui sont détaillés dans la remarque 2 pendant l'implémentation.

Conclusion

Le travail a pu explorer certaines occurrences de « *ma/m'* » dans le corpus à l'aide de 11 hypothèses au total divisées entre les 3 catégories de *ma* : 5 pour l'auxiliaire *ma*, 2 pour le nom *ma* et 4 pour la postposition *ma*.

Chaque hypothèse a été revue en règle quelconque après les remarques d'infirmité de confirmation faites au sujet de leurs occurrences correspondantes. De ce fait, le travail a mis en évidence 11 règles précises sur la reconnaissance automatique et systématique des *ma* dans le corpus d'analyse.

Chacune des 11 règles sont la reformulation de l'hypothèse en tenant compte de leurs pourcentages de confirmation. Les exceptions aux règles sont cependant relatives aux pourcentages d'infirmité des hypothèses.

Dans ce cas, les règles chez l'auxiliaire *ma* peuvent traiter automatiquement et systématiquement les *ma* jusqu'à hauteur de 406/757 (53,63%). Les règles du nom *ma* peuvent traiter jusqu'à hauteur de 39/757 (5,15%). Chez la postposition *ma*, les règles peuvent traiter efficacement 323/757 (42,66) *ma*. Le cumul des chiffres sur les *ma* couvert par les 11 règles s'élève alors à 768/757 (101,45%). Ce nombre (768) dépasse de 11 *ma* le nombre total de *ma* dans le corpus (757). Cet état s'explique par le fait que certaines occurrences de *ma* sont couvertes répétitivement par les mêmes règles. Cela ne peut nuire nullement à la qualité du traitement en ce sens où le programme pourra sauter les occurrences déjà traitées lors de l'application d'une règle qui la précède.

On ne peut pas aussi confirmer que l'ensemble des 757 occurrences de *ma* dans le corpus sont intégralement couvertes par les 768 *ma* explorés par les règles. La répétition dans la couverture des *ma* peut avoir augmenté les chiffres jusqu'à 768 sans pour autant couvrir les 757 « *ma/m'* » du corpus.

En fin de mot, il est devenu évident que la méthode de l'ordre d'occurrence des unités dans le corpus est une solution à la reconnaissance automatique et systématique des occurrences absolues du corpus en général et des occurrences absolues de *ma* en particulier. Cette reconnaissance automatique de la catégorie lexicale des *ma* dans un corpus ouvre la voie vers la reconnaissance automatique des sens des différents *ma*. Ces règles trouvent leur application dans les correcteurs orthographiques et syntaxiques, les traducteurs automatiques et dans bien d'autres mécanismes de traitement artificiel du bamanankan.

Partant de toutes les évidences démontrées dans les exposés sur les 3 *ma*, je suis convaincu qu'une bonne base de l'industrie du traitement virtuelle automatique et systématique est mise à la disposition des ingénieurs et spécialistes de la programmation informatique. A ce stade, plus rien n'empêche quelqu'un à réaliser un traducteur automatique bamanankan-français ou bamanankan-english et ou un correcteur syntaxique bamanan sous la base des acquis, minimales soient-ils, en reconnaissance automatique livrés par la présente contribution. Je recommande fortement à tout programmeur informaticien les acquis du présent travail.

Références bibliographiques

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

Diallo, Youssouf et al., 2006, *Bamanankan maben*, Bamako, Donniya.

Dnafla, 1997, *Bamanankan sariyasun*, Bamako, Dnafla.

Dukure, Mamadu, 2009, *Sebenko bere*, Bamako, Makdas sebenca.

Dukure, Mamadu, Baalo, Isiyaka, 2008, *Danɛgafe Wagadu*, Bamako, Makdas sebenca.

Vydrine, Valentin, 1999, *Mandén-Ankile Danɛgafe*, St Petersburg, Dimitry Bulanin Publishing House.