Données dialectométriques du birifor

KAMBOU Bagboulissan Gilbert
Équipe École Doctorale des Lettres
Sciences Humaines et Communication (E.D/LESHCO)/
Université Joseph KI-ZERBO (Burkina Faso)
bgilbertkambou@gmail.com
&
ASSANVO Amoikon Dyhie
Université Félix Houphouët-Boigny (Côte d'Ivoire)
adyhies@gmail.com

Résumé :
Le birifor, langue gur parlée en Côte d'Ivoire, au Burkina Faso et au Ghana, présente une importante variation dialectale encore peu étudiée de façon systématique. Cette recherche s'appuie sur une approche dialectométrique pour mesurer quantitativement les distances linguistiques entre différentes variétés du birifor. À partir d'un corpus lexical et morphosyntaxique collecté dans plusieurs localités, l'étude applique des méthodes statistiques (comme la distance de Levenshtein) pour analyser les écarts et dégager des tendances dans la structuration dialectale. Les résultats permettent de proposer une cartographie des parlers birifor, de hiérarchiser les proximités linguistiques et de mieux comprendre les dynamiques internes de cette langue. Ce travail contribue à la description fine du birifor et ouvre des perspectives pour des projets de standardisation ou d'aménagement linguistique.

Mots-clés : dialectométrie, birifor, variation linguistique, distance linguistique, cartographie dialectale.

Abstract:
Birifor, a Gur language spoken in Côte d'Ivoire, Burkina Faso, and Ghana, exhibits significant dialectal variation that remains largely underexplored through systematic analysis. This study adopts a dialectometric approach to quantitatively assess linguistic distances between various Birifor varieties. Based on a lexical and morphosyntactic corpus gathered from different locations, statistical methods—such as Levenshtein distance—are applied to analyze variation and identify patterns in the internal structure of the language. The findings lead to a proposed mapping of Birifor dialects, highlighting linguistic proximities and providing deeper insight into the language's internal dynamics. This research contributes to the detailed description of Birifor and offers a foundation for future efforts in linguistic standardization and planning.

Keywords: dialectometry, Birifor, linguistic variation, language distance, dialect mapping.

INTRODUCTION

Le birifor est une langue transfrontalière, parlée au-delà des frontières du Burkina Faso. Il est parlé dans trois pays notamment le nord du Ghana, le nord de la Côte d'Ivoire et le sud-ouest du Burkina Faso qui constitue notre zone d'étude. Le Burkina Faso compte treize (13) régions. Nous pouvons constater l'existence du birifor dans la Bougouriba, le Poni et le Noumbiel, trois (03) provinces sur quatre (04) que compte la région qui a pour chef-lieu le Poni. Nous dénombrons deux cents trois (203) villages exclusivement birifor sans compter les villages mixtes d'où la communauté birifor et autre y vivent. Démographiquement, les birifor représentent 0,7%, soit 143536 habitants de la population burkinabè selon le recensement général de 2019. Il est classé quatorzième parmi les langues les plus parlées d'après ce recensement, et troisième après le dagara et lobiri dans la région du sud-ouest. D'après Lamboni (2017 : 22) : « Le nombre de locuteurs actifs est un indice pertinent

car il contribue à la puissance innée de la langue. Il est davantage pertinent quand les positions géographiques des locuteurs sont diversifiées et stratégiques. ». Le birifor est une langue parlée par les birifor (locuteurs). La sous-commission du birifor a été reconnue officiellement le 04 juin 2010. Le birifor est une langue vernaculaire, c'est-à-dire servant de moyen de communication seulement à l'intérieur de la communauté. On remarque sa vivacité dans ses zones de prédilection. Il est de type gur et est classé dans la famille des langues Niger-Congo, selon Manessy (1975). C'est une langue dite minoritaire, utilisée de plus en plus dans l'alphabétisation, dans les confessions religieuses et parfois dans le tribunal de Grande Instance de Gaoua comme langue d'interprétation. En ce qui concerne l'alphabétisation, il faut noter qu'elle était plus organisée dans les missions catholiques et protestantes dans le but d'aider la communauté birifor à lire les textes sacrés. Pour Gustav Warneck (1834-1910), cité par Lamboni (2017 : 25), l'objectif était : « apprendre à lire pour faciliter l'accès à la Bible et aux Ecritures Saintes, procurer des connaissances qui permettront aux élèves de rester au service des Missions, rassembler des jeunes qui ne sont pas encore atteints par la prédication et les convertir ». En outre, le programme national d'alphabétisation a tenu compte du birifor dans l'éducation non formelle dans ces dernières années. A cet effet, des centres pour adultes ont été ouverts. Il est aussi utilisé par moment par une ONG dénommée Stratégie de Scolarisation Accélérée/Centre Passerelle (SSAP) comme langue médium dans les classes d'alphabétisation pour aider les enfants de neuf (09) à douze (12) ans à s'intégrer dans le système scolaire. Toujours dans la même perspective, en 2018 l'Alliance Biblique du Burkina Faso en partenariat avec SIL Burkina Faso, ont produit un document intitulé ''Je lis et j'écris en birifor''. C'est un guide de transition au profit de ceux qui ont déjà le français comme acquis. Depuis cette date, des lycéens, des universitaires et des fonctionnaires apprennent à lire et à écrire en birifor. Faisons foi à la description ci-dessus, nous tenterons de répondre à ces questions : Quelles sont les principales variantes dialectales du birifor dans les différentes régions où il est parlé ? Quelles différences phonétiques, lexicales et syntaxiques observe-t-on entre ces variantes ? Quelle est la distance dialectométrique entre les parlers identifiés ? Quels facteurs (géographiques, historiques, sociolinguistiques) influencent la variation dialectale du birifor ? Partant de ces questions, nous conjecturons que les variantes du birifor présentent une variation significative corrélée à la distance géographique et aux zones d'influence socioculturelle, ce qui permet une classification dialectométrique structurée. Pour terminer, la présente étude tente de mesurer quantitativement les différences entre variantes à l'aide d'outils dialectométriques.

1- Cadre théorique

Le présent travail de recherche s'inscrit dans le cadre général de la géographie linguistique qui se donne pour mission de répartir les variantes d'une langue dans une région donnée à partir des données d'un corpus en faisant ressortir leurs ressemblances et leurs différences. Autrement dit, localiser les diverses variantes d'une ou plusieurs langues en vue de leur description et de leur comparaison pour dégager leurs traits communs et leurs traits particuliers. Elle renferme plusieurs méthodes dites méthodes dialectométriques. Le terme dialectométrie étant une combinaison bi-disciplinaire regroupant les données de la géographie linguistique et de la taxinomie numérique.

2- Cadre méthodologique

Le birifor parlé au Burkina Faso en rappel peut être identifié dans ces trois (03) provinces du Sud-Ouest (Bougouriba, Poni, Noumbiel) dudit pays. Les localités birifor s'élèvent à deux cent trois (203) hormis les nombreux villages occupés majoritairement par les birifor et les lobi que nous avons volontairement omis, selon DA Inyinibon (2017). Pour la réussite de l'enquête de terrain, nous avons d'ores et déjà balisé le terrain à l'aide de la pré-enquête, qui nous a permis de choisir les dix (10) points d'enquête, voire les dix (10) parlers du birifor sur les 203 localités que compte le terrain d'enquête. Les points d'enquête sont entre autres Bamako, Batié, Batié blé, Goran, Hello, Kpuéré, Malba, Ponalatéon, Sidoumoukar et Talière. Ce test préalable déjà établi, nous a permis de procédé à

la technique de l'entretien qui a consisté à entrer en contact direct avec les informateurs identifiés. A cet effet, un questionnaire est soumis aux informateurs en français pour la traduction. Nous avons opté pour la plupart, des informateurs bilingues (birifor-français). Ils sont issus de divers rangs sociaux, à savoir les cultivateurs/éleveurs, les enseignants, les chefs coutumiers et pasteurs. Le choix est orienté vers ces modestes personnes eu égard aux réalités du thème auquel nous faisons face. Les raisons qui ont motivé à leur choix sont la maitrise des deux langues (français-birifor) et le fait qu'ils sont natifs desdites localités. Ils sont majoritairement du genre masculin. L'âge de ces informateurs est compris entre 35 à 75 ans. Le choix des informateurs joue un rôle déterminant dans tout travail de recherche. Pour la transcription du corpus, nous nous sommes servi de la notation phonique large qui impose l'utilisation des symboles de l'Alphabet Phonétique International (API).

3- Transformation des données dialectales en données numériques

Un extrait de cent (100) mots des cent trente (130) notions fondamentales de Möhlig recueillis dans chaque point d'enquête, nous permet de calculer les pourcentages de ressemblance et les pourcentages de différence linguistique. L'obtention de ces différents pourcentages sert à établir les coefficients de proximité linguistique et les coefficients de distance linguistique. Ces coefficients recueillis nous permettent également de calculer les moyennes de proximité linguistique et les moyennes de distance linguistique des parlers birifor. Il faut noter que les cent (100) mots sont inclus dans les huit cent treize (813) mots qui constituent le corpus de base. Le calcul de l'indice de partition est rendu possible par le biais des moyennes de proximité et de distance linguistique. Les pourcentages de différence et de ressemblance obtenus par paire de localité et par paire de mots à partir du corpus des cents (100) mots, sont mis en annexe.

3-1- Calcul des pourcentages de ressemblances et de différences linguistiques

Les pourcentages de ressemblance et de différence sont mesurés selon un procédé de calcul par comparaison de mots deux à deux. Il s'agit de prendre un mot d'un point d'enquête donné et le comparer au mot de même sens d'un autre point d'enquête. Ce faisant, le même mot en français est soumis et traduit dans les différents dialectes du birifor selon les localités retenues de la zone d'étude. La comparaison des formes du mot obtenues par paire de localité consiste à les observer en vue d'identifier les ressemblances et les différences constatées au niveau des segments tels que les consonnes, les voyelles, et les supra-segments qui sont les tons. Cette comparaison consiste à opposer les formes du singulier entre elles et celles du pluriel entre elles si celles-ci existent. C'est le résultat de cette comparaison qui permet de calculer le pourcentage de ressemblance et de différence. Malgoubri (2011 : 11), s'inspire du mode de calcul de Goebl (1981) intitulé « mode de similarité ou d'identité », pour calculer les pourcentages de ressemblance (PR) et les pourcentages de différence à partir des formules suivantes :

$$PR = \frac{NR\ X\ 100}{NR+ND}$$

PR = pourcentage de ressemblances
ND = nombre de différences
NR = nombre de ressemblances

$$PD = \frac{ND\ X\ 100}{ND+NR}$$

PD = pourcentage de différences
ND = nombre de différences
NR = nombre de ressemblances

Le calcul des pourcentages de ressemblance ou de différence peut s'opérer en suivant le principe de la relation de complémentarité qui spécifie la ressemblance et la différence par : PR + PD = 100 ; le chiffre 100 du fait que nous avons retenu 100 mots pour les calculs de pourcentages en vue d'hiérarchiser l'espace dialectal du birifor. De cette relation, nous pouvons déduire le pourcentage de

différences, le pourcentage de ressemblances étant déjà connu : PD = 100 – PR. Soulignons que cette relation suppose une hiérarchisation contraire en ce sens qu'elle est effectuée à partir des PR ou des PD. Pour la comparaison des notions par paire de localités et par paire de mots, nous appliquons le principe d'alignement et de correspondance. Ce principe consiste à comparer les mots deux à deux sur l'axe paradigmatique en opposant les segments des radicaux et ceux des suffixes. Autrement dit, comparer une consonne à une consonne et une voyelle à une voyelle. Ce principe nous parait le mieux approprié pour le présent travail, vu que le birifor est une langue à classe avec des suffixes de classe divers voire irréguliers tant au niveau du singulier que du pluriel. L'application rigoureuse de ce principe nous éloignerait des pourcentages de ressemblance très réduits ou à des pourcentages de différences exagérés. A titre illustratif, pour obtenir le pourcentage de ressemblance de la notion « langue » réalisée dans la variante de Bamako et celle de Goran, nous avons pour le cas de l'application du principe de non alignement ce qui suit :

Localité	Singulier	Pluriel
Bamako	délĩ́bí-r	délĩ́bí-é
Goran	ɟã́bí-r	ɟã́bí-é

singulier	d	ɛ́	l	ĩ́	b	í	r	Total
	ɟ	ã́	b	í	r			
	0	2	0	2	0	0	0	4

pluriel	d	ɛ́	l	ĩ́	b	í	é	Total
	ɟ	ã́	b	í	é			
	0	2	0	2	0	0	0	4

En ce qui concerne le singulier, nous avons quatre éléments identiques (4 tons) et dix-sept éléments en présence (7 consonnes, 5 voyelles et 5 tons).

$$PR = \frac{4\ X\ 100}{17}$$
$$PR = 23{,}529$$

Pour le pluriel, nous avons quatre éléments identiques (4 tons) et dix-neuf éléments en présence (5 consonnes, 7 voyelles et 7 tons).

$$PR = \frac{4\ X\ 100}{19}$$
$$PR = 21{,}052$$

Comme déjà souligné ci-dessus pour le calcul des pourcentages, nous additionnons le pourcentage de ressemblance de la forme au singulier et celui de la forme au pluriel divisés par 2. Dans le cas présent de la notion « langue », il s'agit de $\frac{23{,}529+21{,}052}{2}$; ce qui nous donne un pourcentage de **22,290**% que fait montrer le principe de non alignement appliqué à la notion « langue » à Bamako et à Goran. Nous remarquons un faible taux de ressemblance tandis que phonologiquement le nombre d'éléments identiques (bír) est visiblement appréciable. Toutefois, l'application du principe d'alignement donne ce qui suit:

singulier	d	ɛ́	l	ĩ́	b	í	r	Total
			ɟ	ã́	b	í	r	
	0	0	0	2	2	4	2	10

pluriel	d	ɛ́	l	ĩ́	b	í	é	Total
			ɟ	ã́	b	í	é	
	0	0	0	2	2	4	4	12

Pour le singulier, le nombre d'éléments identiques est dix (4 consonnes, 2 voyelles et 4 tons) et dix-sept éléments en présence (7 consonnes, 5 voyelles et 5 tons).

$$PR = \frac{10\ X\ 100}{17}$$

$$PR = 58{,}823$$

Pour le pluriel, il y a dix-neuf éléments en présence dont 5 consonnes, 7 voyelles et 7 tons. Les éléments identiques sont au nombre de douze dont 2 consonnes, 4 voyelles et 6 tons.

$$PR = \frac{12\ X\ 100}{19}$$

$$PR = 63{,}157$$

Nous faisons le cumule des deux pourcentages et le résultat divisé par 2. Ce qui donne pour la notion « langue » à Bamako et à Goran un pourcentage de :

$$PR = \frac{58{,}823+63{,}157}{2} = \mathbf{60{,}99}$$

Ce taux de ressemblance est nettement plus proche de la réalité dû au principe d'alignement qui consiste à opposer radical au radical et suffixe au suffixe, le préfixe étant inexistant en birifor. Ce principe appliqué pour l'opération de cette notion a permis de voir un rapprochement plus raisonnable en termes de ressemblance et de réduire considérablement cette distance exagérée. C'est ce qui témoigne la nécessité du principe d'alignement dans le présent travail. En général, le principe d'alignement appliqué avec rigueur nous épargne des distances à la limite douteuses et permet de réaliser un travail qualitatif.

3-2- Calcul des pourcentages de ressemblance (PR) des localités

Pour calculer le pourcentage de ressemblance d'une notion traduite par paire de localités, nous opposons les formes du singulier entre elles et les formes du pluriel entre elles également si elles existent. Ainsi, le nombre de sons identiques des formes comparées sont multipliés par 100 et le résultat divisé par le nombre total d'éléments en présence. Lorsque c'est un mot à double forme (singulier et pluriel), le pourcentage de la forme du singulier et celui de la forme du pluriel obtenus, sont additionnés et divisés par 2. Toutefois, pour obtenir le pourcentage de ressemblance d'un mot entre deux localités dont l'une des localités a une seule forme (singulier), le nombre d'éléments identiques des formes du singulier sont multipliés par 100 et le résultat divisé par le nombre total d'éléments en présence c'est-à-dire les formes du singulier plus la forme du pluriel. Pour le calcul de pourcentage de ressemblances, Malgoubri (2011) s'est servi de la formule suivante:

$$PR = \frac{NR\ X\ 100}{NR+ND}$$

PR = pourcentage de ressemblances
NR = nombre de ressemblances
ND = nombre de différences

Nous calculons à titre d'exemple le pourcentage de ressemblances (PR) de la notion « mari » à Batié blé et à Hello.

Singulier pluriel

Batié blé	sír-ɛ́	sír-bɛ́
Hello	sír-á	sír-bá

Pour le singulier, nous enregistrons 12 éléments en présence dont 8 segments et 4 tons. Le nombre d'éléments identiques est 10 (4 consonnes, 2 voyelles et 4 tons).

$$PR = \frac{10\ X\ 100}{12} = 83{,}333$$

Pour ce qui est du pluriel, nous dénombrons 10 segments et 4 tons dont 14 éléments en présence. Le nombre d'éléments identiques s'élève à 12 soient 6 consonnes, 2 voyelles et 4 tons.

$$PR = \frac{12\ X\ 100}{14} = 85{,}714$$

En définitive, le pourcentage de ressemblances de la notion « mari » des deux localités est :

$$PR = \frac{83{,}333+85{,}714}{2}$$

$$PR = \mathbf{84{,}52}$$

3-3- Calcul des pourcentages de différences (PD) des localités

Pour ces calculs, Malgoubri (op.cit.) a utilisé cette formule :

$PD = \frac{ND\ X\ 100}{ND+NR}$

PD = pourcentage de différences
NR = nombre de ressemblances
ND = nombre de différences

Le pourcentage de ressemblance (PR) étant déjà connu, en nous basant sur le principe de la complémentarité entre la ressemblance et la différence qui donne la formule suivante : PR + PD = 100, nous avons calculé le pourcentage de différences (PD) sous cette formule : PD = 100 – PR. Sous cet angle, sachant que les pourcentages de ressemblances (PR) de la notion « mari » des deux localités au singulier est égal à **83,33%** et au pluriel **85,71%**, nous avons :

Pour le pourcentage de différence (PD) au singulier (100 – 83,33%) = 16,66%, et pour celui du pluriel (100 – 85,71%) = 14,28%. Toujours dans cette perspective, le pourcentage de différences définitif est égal à :

PD = 100 - 84,523 % = **15,47%**

C'est le même procédé que nous avons observé pour tous les calculs de pourcentages de ressemblances et de différences par paire de localités et par paire de mots des dix localités retenues pour notre zone d'étude.

3-4- Calcul des coefficients de proximité et de distance linguistiques

Les coefficients de proximité linguistique et les coefficients de distance linguistique sont des valeurs chiffrées qui caractérisent les points d'enquête par paire de localités en termes de ressemblance et de différence. Les coefficients de proximité et de différence sont obtenus en faisant la somme de tous les pourcentages de ressemblances et de différences établis par bipoints divisée par le nombre total de pourcentages. Les coefficients de proximité et de distance linguistiques s'obtiennent par les formules ci-dessous :

$CPL = \frac{1}{N}\sum_{i=1}^{n}(PR)i$

i = succession de 1 à n

PR = pourcentage de ressemblances
N = nombre de pourcentage de ressemblances
$\sum = somme\ de\ \ldots$

$CDL = \frac{1}{N}\sum_{i=1}^{n}(PD)i$

i = succession de 1 à n

PD = pourcentage de différences
N = nombre de pourcentage de différences
$\sum = somme\ de\ \ldots$

3-5- Calcul des coefficients de proximité linguistique

Pour obtenir le coefficient de proximité linguistique des localités, nous additionnons les différents pourcentages de ressemblances par bipoint. La somme des pourcentages de ressemblances obtenue est ensuite divisée par 100. En plus de la formule citée ci-dessus, le coefficient de proximité linguistique peut s'obtenir également par le biais de la formule simplifiée suivante :

$\mathrm{CPL} = \frac{1}{\mathrm{N}}(\mathrm{X}_1 + X_2 + X_3 + \ldots X_{100})$ N = nombre total de pourcentages : 100

X_1 = premier pourcentage

X_{100} = dernier pourcentage

A titre illustratif du bipoint Talière/Goran, nous avons :

$$\mathrm{CPL} = \frac{\text{Somme des PR}}{100}$$

La somme des pourcentages de ressemblances (PR) est égale à 6788,198 En appliquant la formule, cela donne :

$$\mathrm{CPL} = \frac{6788{,}198}{100} = 67{,}88$$

En définitive, les 67,88 entre la localité Talière et celle de Goran est le chiffre exprimé en pourcentage en termes de coefficient de proximité linguistique (CPL) desdites localités. La même méthode est observée pour le calcul de tous les coefficients de proximité linguistique (CPL). Nous présentons dans le tableau ci-dessous les coefficients de proximité linguistique (CPL) des dix (10) paires de localités.

Tableau 1 : coefficients de proximité linguistique

Les paires de localités	Les coefficients de proximité linguistique
Bam/Bat	66,36
Bam/Bab	66,91
Bam/Gor	65,24
Bam/Hel	77,38
Bam/Kpué	61,75
Bam/Mal	81,33
Bam/Pon	70,90
Bam/Sid	81,28
Bam/Tal	80,27
Bat/Bab	81,28
Bat/Gor	75,89
Bat/Hel	78,41
Bat/Kpué	77,40
Bat/Mal	73
Bat/Pon	79,37
Bat/Sid	69,76
Bat/Tal	66,25
Bab/Gor	77,30
Bab/Hel	72,16
Bab/Kpué	75,59
Bab/Mal	73,28
Bab/Pon	77,27
Bab/Sid	73,14
Bab/Tal	67,91
Gor/Hel	71,84
Gor/Kpué	77,25
Gor/Mal	73,07

Gor/Pon	83,45
Gor/Sid	71,43
Gor/Tal	67,88
Hel/Kpué	64,75
Hel/Mal	83,04
Hel/Pon	72,03
Hel/Sid	84,29
Hel/Tal	79,74
Kpué/Mal	67,52
Kpué/Pon	77,34
Kpué/Sid	67,97
Kpué/Tal	63,18
Mal/Pon	76,64
Mal/Sid	85,40
Mal/Tal	82,13
Pon/Sid	74,33
Pon/Tal	71,69
Sid/Tal	81,22

Sources : enquête menée par KAMBOU (2024)

Valeurs des symboles et abréviations :

Bam = Bamako ; Bat = Batié ; Bab = Batié blé ; Gor = Goran ; Hel = Hello ; Kpué = Kpuéré ; Mal = Malba ; Pon = Ponalatéon ; Sid = Sidoumoukar ; Tal = Talière.

3-6- Calcul des coefficients de distance linguistique

La valeur des coefficients de proximité linguistique (CPL) étant déjà connue, nous sommes allé sur la base de la relation de complémentarité entre les coefficients de proximité linguistique (CPL) et les coefficients de distance linguistique (CDL) qui donne cette formule : CPL + CDL = 100. Ainsi, nous avons la formule ci-après : CDL = 100 – CPL. En prenant par exemple la paire de localités Talière/Goran, leur coefficient de distance linguistique a été obtenu en faisant : 100 – 67,881 = 32,11%. Sachant que les CPL et les CDL des localités sont déjà calculés (voir annexe), nous présentons ici les résultats des coefficients de distance linguistique par paire de localités.

Tableau 2 : coefficients de distance linguistique

Les paires de localités	Les coefficients de distance linguistique
Bam/Bat	33,64
Bam/Bab	33,09
Bam/Gor	34,76
Bam/Hel	22,62
Bam/Kpué	38,25
Bam/Mal	18,67
Bam/Pon	29,10
Bam/Sid	18,72
Bam/Tal	19,73
Bat/Bab	18,72
Bat/Gor	24,11
Bat/Hel	21,59
Bat/Kpué	22,60
Bat/Mal	27
Bat/Pon	20,63

Bat/Sid	30,24
Bat/Tal	33,75
Bab/Gor	22,70
Bab/Hel	27,84
Bab/Kpué	24,41
Bab/Mal	26,72
Bab/Pon	22,73
Bab/Sid	26,86
Bab/Tal	32,09
Gor/Hel	28,16
Gor/Kpué	22,75
Gor/Mal	26,93
Gor/Pon	16,56
Gor/Sid	28,57
Gor/Tal	32,12
Hel/Kpué	35,25
Hel/Mal	16,96
Hel/Pon	27,97
Hel/Sid	15,71
Hel/Tal	20,26
Kpué/Mal	32,48
Kpué/Pon	22,66
Kpué/Sid	32,03
Kpué/Tal	36,82
Mal/Pon	23,36
Mal/Sid	14,60
Mal/Tal	17,87
Pon/Sid	25,67
Pon/Tal	28,31
Sid/Tal	18,78

Sources : enquête menée par KAMBOU (2024)

4. Population cible et valeur corrigée

4-1 Population cible

Pour Grawitz (1998 : 593), la population cible est perçue comme un ensemble des membres d'un groupe spécifique sur lequel les résultats seront applicables. C'est dans cette optique nous stipulons que la population cible dans la présente recherche est l'ensemble de tous les locuteurs birifor situés dans les différentes localités ou villages de la zone d'étude. Cette population cible est localisée dans le Sud-Ouest du Burkina Faso plus précisément dans trois provinces sur quatre que compte la région. Cette zone s'étend de Diébougou jusqu'à Kpuéré, en passant par Gaoua et Batié. La zone d'étude compte 203 localités ou villages. A cet effet, nous visons dans ce présent travail à mettre en exergue les différentes variantes du birifor sur le plan dialectologique et dialectométrique.

4-2- Valeur corrigée

Il faut noter que les données de l'enquête n'ont pas été collectées dans tous les villages de la zone d'étude. Nous avons mené l'enquête par zone, selon le regroupement des birifor. Et à chaque zone nous enregistrons exclusivement les différents parlers (localités) qui ont été identifiés. C'est fort de ce constat, avec ces coefficients obtenus nous avons eu recours à la valeur corrigée, car les données collectées n'ont pas couvert toutes les localités de la zone d'étude. Cette expression valeur corrigée émane de Guiter (1973) pour corriger les pourcentages de ressemblance et de différence qui

caractérisent les points d'enquête par paire de localités. C'est dans ce sens que Guiter parle de « d'atlas clairsemé » ou de « densité clairsemée » par opposition à « l'atlas exhaustif » ou « densité exhaustive ». De ce point de vue, il stipule que l'atlas clairsemé est le nombre de points étudiés dans une zone d'étude donnée tandis que l'atlas exhaustif englobe toutes les localités ou villages de la zone d'étude. Pour combler les insuffisances qui pourraient compromettre les données réelles qui existent en termes de ressemblance et de différence, il propose la formule ci-après :

$$\Delta N = N\sqrt{\frac{(DO-D)}{D}}\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

ΔN: variation de pourcentages des différences (proximité ou distance linguistique)
N : pourcentage de différences ou de ressemblances entre deux points contigus,
Do : densité exhaustive (nombre total de localité)
D : densité clairsemée (nombre de localités étudiées)

Notre zone de recherche compte 203 villages ou localités. Parmi celles-ci, nous avons retenu 10 localités ou points d'enquête après l'étape de la pré-enquête. Les 10 localités symbolisent la densité clairsemée, tandis que les 203 localités déterminent la densité exhaustive. Ainsi, nous avons calculé la variation de pourcentages ΔN.

4-3- Variation de pourcentages ΔN de la zone d'étude

Nous caractérisons la valeur de pourcentages ΔN de l'espace d'étude dans le but de procéder à la correction des coefficients de proximité linguistique (CPL) et des coefficients de distance linguistique. Soit :
D0 : 203 (nombre total de villages)
D : 10 (nombre de villages où les données ont été collectées)

$$\Delta N = N\sqrt{\frac{(DO-D)}{D}}\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

$$\Delta N = N\sqrt{\frac{(203-10)}{10}}\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

$$\Delta N = N\sqrt{\frac{(193)}{10}}\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

$$\Delta N = N\sqrt{19{,}3}\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

$$\Delta N = N\,(4{,}39)\left(\frac{N}{100}\right)^2\left(\frac{100-N}{100}\right)^2$$

4-3-1- Calcul des coefficients de proximité linguistique pondérés (CPLP) et des coefficients de distance linguistique pondérés (CDLP)

La valeur de ΔN de la zone d'étude étant déjà connu, nous avons déterminé ΔN de chaque paire de $_1$localités avant de calculer leurs coefficients de proximité linguistique pondérés (CPLP) et leurs coefficients de distance linguistique pondérés (CDLP). Pour les dix (10) variétés, nous dénombrons 45 coefficients de proximité linguistique ou 45 coefficients de distance linguistique. Sur ce, les valeurs ΔN des paires de localités sont numérotées de ΔN1 à ΔN45. A titre illustratif, nous calculons ΔN1 du bipoint Sid/Kpué:
Soit N_1 le CPL de Sid/Kpué
$N_1 = 67{,}97$

$$\Delta N1 = (67{,}97)\,(4{,}39)\left(\frac{67{,}97}{100}\right)^2\left(\frac{100-67{,}97}{100}\right)^2$$

$$\Delta N1 = (298{,}38)\left(\frac{67{,}97}{100}\right)^2\left(\frac{100-67{,}97}{100}\right)^2$$

$$\Delta N1 = (298{,}38)\,(0{,}6797)^2\left(\frac{32{,}04}{100}\right)^2$$

$$\Delta N1 = (298{,}38)\,(0{,}6797)^2\,(0{,}3203)^2$$

$\Delta N1 = (298{,}38)\ (0{,}461)\ (0{,}102)$
$\Delta N1 = 14{,}03$

Le coefficient de proximité linguistique pondéré s'obtient en ajoutant ΔN à N ; N étant le coefficient de proximité linguistique non pondéré. Sur ce, nous avons la formule ci-après :

CPLP = N + ΔN

Pour le cas de Sid/Kpué à titre d'exemple, nous avons : N1= 67,96 et $\Delta N1 = 14{,}02$. Donc CPLP = 67,97 + 14,03
CPLP de Sid/Kpué = 82

En ce qui concerne le coefficient de distance linguistique pondéré, il s'obtient en ôtant la valeur de ΔN de N′ ; N'étant la valeur du coefficient de distance linguistique non pondéré. A cet effet, la formule suivante s'applique à cette opération :

CDLP = N' - ΔN

Sachant que N' = 32,03 et $\Delta N = 14{,}03$
CDLP = 32,03 – 14,03
CDLP de Sid/Kpué = 18

Nous avons procédé de la même manière pour calculer les coefficients de proximité linguistique pondérés et les coefficients de distance linguistique pondérés des 45 bipoints.

Tableau 3 : CPLP et CDLP des paires de localités

Les paires de localités	ΔN	Coefficients de proximité linguistique pondérés	Coefficients de distance linguistique pondérés
Bam/Bat	14,48	80,84	19,16
Bam/Bab	14,31	81,22	18,78
Bam/Gor	14,60	79,84	20,16
Bam/Hel	10,35	87,73	12,27
Bam/Kpué	15,07	76,82	23,18
Bam/Mal	8,02	89,35	10,65
Bam/Pon	13,12	84,02	15,98
Bam/Sid	8,24	89,52	10,48
Bam/Tal	8,62	88,89	11,11
Bat/Bab	8,24	89,52	10,48
Bat/Gor	11,11	87	13
Bat/Hel	9,72	88,13	11,87
Bat/Kpué	10,37	87,77	12,23
Bat/Mal	12,27	85,27	14,73
Bat/Pon	9,20	88,57	11,43
Bat/Sid	13,54	83,30	16,70
Bat/Tal	14,39	80,64	19,36
Bab/Gor	10,33	87,63	12,37
Bab/Hel	12,68	84,84	15,16
Bab/Kpué	11,17	86,76	13,24
Bab/Mal	12,24	85,52	14,48
Bab/Pon	10,32	87,59	12,41
Bab/Sid	12,34	85,48	14,52
Bab/Tal	14,01	81,92	18,08
Gor/Hel	12,85	84,69	15,31

Gor/Kpué	10,30	87,55	12,45
Gor/Mal	12,30	85,37	14,63
Gor/Pon	6,88	90,33	9,68
Gor/Sid	12,95	84,38	15,62
Gor/Tal	14,11	81,99	18,01
Hel/Kpué	14,76	79,51	20,49
Hel/Mal	7,03	90,07	9,93
Hel/Pon	12,77	84,80	15,20
Hel/Sid	6,30	90,59	9,41
Hel/Tal	9,11	88,85	11,15
Kpué/Mal	14,16	81,68	18,32
Kpué/Pon	10,35	87,69	12,31
Kpué/Sid	14,03	82	18
Kpué/Tal	14,93	78,11	21,89
Mal/Pon	10,66	87,30	12,70
Mal/Sid	5,73	91,13	8,87
Mal/Tal	7,53	89,66	10,34
Pon/Sid	11,70	86,03	13,97
Pon/Tal	12,91	84,60	15,40
Sid/Tal	8,22	89,44	10,56

Sources : enquête menée par KAMBOU (2024)

A l'issue de ces différentes opérations qui nous ont conduit à ces pourcentages ci-dessus, nous pouvons à présent établir les matrices des données et également calculer les moyennes de proximité linguistique (MPL) et les moyennes de distance linguistique (MDL) des localités. La matrice se définit comme un tableau orthogonal dans lequel sont consignées les ressemblances et les différences en termes de pourcentage obtenu par confrontation entre deux localités et par paire de mots, selon Goebl (1983). Les matrices sont susceptibles d'être présentées de deux manières, celle à double entrée à forme rectangulaire ou carrée et celle simplifiée à forme orthogonale. La forme simplifiée est beaucoup plus usitée par les dialectométriciens que celle à double entrée.

a.

b. Matrice à double entrée

	A	B	C	D
A	-	a	b	d
B	a	-	c	e
C	b	c	-	f
D	d	e	f	-

Figure nº : matrice complexe

b. Matrice simplifiée

	A			
B	a	B		
C	b	c	C	
D	d	e	f	D

Figure nº : matrice simplifiée

4-3-2 Matrice des coefficients de proximité linguistique pondérés

Cette matrice est composée des coefficients de proximité linguistique pondérés des 10 localités par bipoint.

	Bam			
Bat	80,84	Bat		
Bab	81,22	89,52	Bab	
Gor	79,84	87	87,63	Gor

Hel	87,73	88,13	84,84	84,69	Hel					
Kpué	76,82	87,77	86,76	87,55	79,51	Kpué				
Mal	89,35	85,27	85,52	85,37	90,07	81,68	Mal			
Pon	84,02	88,57	87,59	90,33	84,80	87,69	87,30	Pon		
Sid	89,52	83,30	85,48	84,38	90,59	82	91,13	86,03	Sid	
Tal	88,89	80,64	81,92	81,99	88,85	78,11	89,66	84,60	89,44	Tal

Matrice des coefficients de proximité linguistique pondérés

La matrice se lit horizontalement jusqu'au croisement de la localité et puis se lit verticalement. Prenons par exemple les coefficients liés au point d'enquête Gor : 79,84 ; 87 ; 87,63 ; 84,69 ; 87,55 ; 85,37 ; 90,33 ; 84,38 ; 81,99. Toutefois, les coefficients liés au point d'enquête Bam se lisent verticalement et ceux liés au point d'enquête Tal se lisent horizontalement.

4-3-3- Matrice des coefficients de distance linguistique pondérés

Cette matrice est constituée de coefficients de distance linguistique pondérés des 10 localités par paire de localités.

	Bam									
Bat	19,16	Bat								
Bab	18,78	10,48	Bab							
Gor	20,16	13	12,37	Gor						
Hel	12,27	11,87	15,16	15,31	Hel					
Kpué	23,18	12,23	13,24	12,45	20,49	Kpué				
Mal	10,65	14,73	14,48	14,63	9,93	18,32	Mal			
Pon	15,98	11,43	12,41	9,67	15,20	12,31	12,70	Pon		
Sid	10,48	16,70	14,52	15,62	9,41	18	8,87	13,97	Sid	
Tal	11,11	19,36	18,08	18,01	11,15	21,89	10,34	15,40	10,56	Tal

Matrice des coefficients de distance linguistique pondérés

4-3-4 Calcul des moyennes de proximité et des moyennes de distance linguistique

En nous appuyant sur les coefficients préalablement calculés, les moyennes de proximité linguistique (MPL) et les moyennes de distance linguistique (MDL) s'obtiennent en faisant la sommation de tous les coefficients de proximité et de distance linguistique pondérés, divisée par le nombre de coefficients qui composent les localités par paire. Pour le calcul de ces moyennes, nous appliquons la formule ci-après préconisée par Malgoubri (1988 : 38) :

$X = \frac{1}{N} \sum_{i=1}^{n} xi$ ou $X = \frac{X1+X2+X3+X4...Xn}{N}$

i= indique une succession de 1 à n

x= coefficient de proximité ou de distance linguistique

N= nombre de coefficients de proximité ou de distance linguistique

X= moyenne de proximité ou de distance linguistique

$\sum$ = somme de

4-3-5- Moyennes de proximité linguistique de chaque localité

A titre illustratif, pour le calcul Kpué, nous avons additionné les coefficients de proximité des localités reliant Kpué de façon horizontale jusqu'à l'intercession de ladite localité puis verticalement. Le résultat de cette addition est divisé par le nombre de combinaison avec la localité Kpué. En ce faisant, nous avons les combinaisons suivantes :

Bam/Kpué = 76,82

Bat/Kpué = 87,77

Bab/Kpué = 86,76

Gor/Kpué = 87,55
Hel/Kpué = 79,51
Mal/Kpué = 81,68
Pon/Kpué = 87,69
Sid/Kpué = 82
Tal/Kpué = 78,11
MPL de Kpué est :

$$X = \frac{76,82+87,77+86,76+87,55+79,51+81,68+87,69+82+78,11}{9}$$

$$X = 83,09$$

Les moyennes de proximité linguistique (MPL) sont portées sous le tableau, à la dernière combinaison de chaque localité.

	Bam									
Bat	80,84	Bat								
Bab	81,22	89,52	Bab							
Gor	79,84	87	87,63	Gor						
Hel	87,73	88,13	84,84	84,69	Hel					
Kpué	76,82	87,77	86,76	87,55	79,51	Kpué				
Mal	89,35	85,27	85,52	85,37	90,07	81,68	Mal			
Pon	84,02	88,57	87,59	90,33	84,80	87,69	87,30	Pon		
Sid	89,52	83,30	85,48	84,38	90,59	82	91,13	86,03	Sid	
Tal	88,89	80,64	81,92	81,99	88,85	78,11	89,66	84,60	89,44	Tal
MPL	**84,24**	**85,67**	**85,60**	**85,42**	**86,57**	**83,09**	**87,26**	**86,77**	**86,87**	**84,90**

Les moyennes de proximité linguistique (MPL) des 10 localités se présentent comme suit :
Bam = 84,24
Bat = 85,67
Bab = 85,60
Gor = 85,42
Hel = 86,57
Kpué = 83,09
Mal = 87,26
Pon = 86,77
Sid = 86,87
Tal = 84,9

4-3-6 Moyennes de distance linguistique de chaque localité

Nous avons observé les mêmes procédés pour les calculs des moyennes de distance linguistique (MDL). Si nous prenons toujours l'exemple de Kpué, nous avons les combinaisons suivantes :

Bam/Kpué = 23,18
Bat/Kpué = 12,23
Bab/Kpué = 13,24
Gor/Kpué = 12,45
Hel/Kpué = 20,49
Mal/Kpué =18,32
Pon/Kpué = 12,31
Sid/Kpué = 18
Tal/Kpué = 21,89
MDL de Kpué est :

$$X = \frac{23,18+12,23+13,24+12,45+20,49+18,32+12,31+18+21,89}{9}$$

$$X = 16,90$$

Les moyennes de distance linguistique (MDL) sont portées sous le tableau, à la dernière combinaison de chaque localité.

	Bam									
Bat	19,16	Bat								
Bab	18,78	10,48	Bab							
Gor	20,16	13	12,37	Gor						
Hel	12,27	11,87	15,16	15,31	Hel					
Kpué	23,18	12,23	13,24	12,45	20,49	Kpué				
Mal	10,65	14,73	14,48	14,63	9,93	18,32	Mal			
Pon	15,98	11,43	12,41	9,67	15,20	12,31	12,70	Pon		
Sid	10,48	16,70	14,52	15,62	9,41	18	8,87	13,97	Sid	
Tal	11,11	19,36	18,08	18,01	11,15	21,89	10,34	15,40	10,56	Tal
MDL	**15,75**	**14,32**	**14,39**	**14,58**	**13,42**	**16,90**	**12,73**	**13,23**	**13,12**	**15,10**

Nous avons pour les dix localités, les moyennes de distance linguistique suivantes :
Bam = 15,75
Bat = 14,32
Bab = 14,39
Gor = 14,58
Hel = 13,42
Kpué = 16,90
Mal = 12,73
Pon = 13,23
Sid = 13,12
Tal = 15,10

CONCLUSION

Il faut noter que l'opération des moyennes de proximité et de distance linguistique nous a permis de constater que la localité de Mal=87,26 a la plus grande moyenne de proximité linguistique, tandis que Kpué=83,09 ressort avec la plus faible moyenne de proximité linguistique. En sus, la localité de Kpué=16,90 obtient la plus grande moyenne de distance linguistique, contrairement à Mal=12,73 qui enregistre la plus faible moyenne de distance linguistique. Nous remarquons également que le rang qu'occupent ces pourcentages des localités Mal et Kpué au niveau des moyennes de proximité est le contraire de celui qu'ils occupent quant aux moyennes de distance linguistique.

Références bibliographiques

Bole-Richard, R. (1982). Langues et dialectes voltaïques : Aperçus dialectologiques.

Chambers, J.K. & Trudgill, P. (1998). Dialectology.

Goebl, H. (2008). Dialectometry: Principes et application.

Goebl, H. (1981). « Eléments d'analyse dialectométrique (avec application à l'AIS) », *Revue de linguistique romane*, 45, 349–420

Goebl, H. (1982). « Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie », in Denkschriften [Mémoires] der Österreichischen Akademie der Wissenschaften, philosophisch-historische Klasse, vol. 157,Vienne, 1–123

Lafage, S. (1993). Les langues voltaïques de Côte d'Ivoire.
Malgoubri, P. (2014). Diversité dialectale et choix d'un dialecte de référence pour un développement durable. Gwenaëlle FABRE, Anne FOURNIER, Lamine SANOGO. Regards scientifiques croisés sur le changement global et le développement - Langue, environnement, culture : Actes du Colloque international de Ouagadougou (8-10 mars 2012), 23-38. https://hal.science/hal-00939890v1
Malgoubri, P. (2011). *Recherches dialectologiques et dialectométriques nuni*, Université de Leiden, Thèse de Ph.D
Malgoubri, P. (2010). Les dialectes mooré: classes nominales et verbes monosyllabiques, *Lomé, Annales de l'Université de Lomé, Tome XXX-1, 257-267*
Nerbonne, J. (2009). Data-Driven Dialectology.
Séguy, J. (1971). La dialectométrie dans l'Atlas linguistique de la Gascogne.