

# *Analyse du corpus*

---

Comme je l'ai déjà mentionné en introduction, cette étude s'appuie sur une analyse quantitative des cooccurrences segmentales et supra-segmentales autorisées par les reconstructions. Je vais présenter la procédure de travail mise en place pour relever les fréquences et les analyser d'un point de vue quantitatif, puis je préciserais quelles cooccurrences ont été étudiées et je terminerai ce chapitre par un index des tableaux.

## ***1 - Analyse quantitative***

Les fréquences d'occurrences servant de base au travail qui va suivre ont été extraites des sources suivantes : - Reconstructions de Meeussen contenues dans B.L.R. (version dactylographiée), - Reconstructions de Guthrie nommées PB X et publiées dans l'appendice 8/1, - Séries comparatives du bantou commun, - Séries partielles du bantou commun.

Toutes ces données<sup>3</sup> ont été informatisées sous file maker, logiciel de tri et à partir de ces fichiers les comptages ont été effectués.

### **1.1 - Traitement des fréquences**

Une fois les fréquences établies, la première étape de l'analyse a été de les distinguer en deux grands types : fréquences attendues vs inattendues.

- Les fréquences - élevées, ou faibles, ou absentes - ont été qualifiées d'**attendues** lorsque leur quantité était en lien avec celle de chaque phonème concerné par la cooccurrence.

---

<sup>3</sup> B.L.R.1 a été informatisé à Berkeley dans le cadre du projet CBOLD. Les séries comparatives et partielles ont été informatisées au laboratoire DDL par Joël Brognart et le PB X par Gisèle Teil-Dautrey.

- Les fréquences - élevées, ou faibles, ou absentes - ont été qualifiées d'**inattendues**, lorsque leur quantité était disproportionnée par rapport à celle de chaque phonème.

Dans le cas des fréquences inattendues, dont le déséquilibre était dû à un excès, j'ai observé chacun des étymons concerné pour identifier la cause de ce déséquilibre. Cette démarche était nécessaire, particulièrement pour les Cs, puisqu'elles peuvent voir une entrée multipliée, du simple fait d'une variation (cf. section 1.1.1, chapitre précédent de cette partie). Cela a pour conséquence d'augmenter artificiellement une cooccurrence donnée. Une fois cet aspect éliminé, j'ai comparé les corpus entre eux, dans le but de savoir si l'excès était redondant de l'un à l'autre. Enfin, la dernière étape a été de confronter ces étymons aux informations contenues dans B.L.R.3 pour valider ou invalider ces étymons et prendre en compte leur niveau de profondeur. Par ailleurs, j'ai régulièrement élargi la recherche à l'ensemble des informations de B.L.R.3 à propos de la distribution suspectée, cette démarche étant bien entendu régulièrement suivie dans le cadre des déséquilibres par défaut.

## 1.2 -Test statistique

Parallèlement, j'ai eu plusieurs fois recours à un test statistique. Le test de chi-2 a été retenu puisqu'il compare des distributions, c'est-à-dire qu'il compare une observation réelle à une observation théorique et il en déduit l'existence ou non d'une relation de dépendance. Ce test offrait l'avantage de pondérer les fréquences observées en fonction du total ligne et du total colonne de chaque segment. Le chi-2 n'a été utilisé que ponctuellement, en particulier pour les cooccurrences consonantiques, dans les cas de regroupement de segments.

## 1.3 - Présentation des données

### • *Tableaux*

Les occurrences et les cooccurrences sont présentées sous forme de tableau de pourcentages, les occurrences ayant permis le calcul des pourcentages se trouvent en annexe. Généralement, au cours du travail, les commentaires **précèdent** le tableau de données.

Dans le but de faire apparaître le plus clairement possible les fréquences inattendues, j'ai organisé les tableaux en tenant compte de la fréquence de chaque segment. Les segments sont classés en allant du plus fréquent au moins fréquent, ainsi je m'attends à trouver les occurrences les plus élevées dans la partie supérieure gauche du tableau. Toute occurrence faible ou absente dans cette partie sera considérée comme suspecte, il en sera de même pour une occurrence élevée

en dehors de cette zone. Cette organisation est surtout intéressante pour les tableaux constitués avec un grand nombre de segments, ce qui sera le cas des cooccurrences incluant les consonnes. Cependant, un ordonnancement par type de segment a aussi sa pertinence, raison pour laquelle le lecteur pourra trouver en annexe le tableau de la cooccurrence étudiée présentée de façon plus conforme à l'usage.

- **Numérotation des tableaux**

Les tableaux sont numérotés selon leur ordre d'apparition dans le corps de la thèse. Chaque **changement dans l'objet étudié** entraîne un changement de numéro et ce changement peut avoir plusieurs sources : - soit changement de corpus (B.L.R.1, Cs etc ...), - soit changement de suites pour un même corpus, - soit changement dans la cooccurrence étudiée.

Les tableaux qui décrivent le **même objet** mais qui le présentent sous une forme différente (en nombre d'occurrences ou en pourcentage ou avec un ordonnancement différent des segments), ces tableaux-là conservent le même numéro, quelle que soit leur place dans la thèse, dans le texte ou en annexe.

#### **1.4 - Comptage du bantou commun**

Avant de commencer l'étude du corpus, je vais préciser les décisions que j'ai dû adopter pour compter les formes à astérisque du bantou commun. Sous le même numéro de série, Guthrie regroupe parfois plusieurs formes lexicales avec ou sans indexation complémentaire. Ces différentes "starred forms" présentent entre elles des variantes qui peuvent être de type vocalique (cf. Cs 1646 ou 1086), ou consonantique (cf. Cs 642 ou 1745 ou...), ou tonal (cf. Cs 674 ou 2061), ou sémantique (cf. Cs 529 ou 1526 ou...). Si les variantes sémantiques ne nous concernent pas directement dans ce travail, par contre les variantes segmentales et suprasegmentales en font partie intégrante. Comment les prendre en compte sans falsifier les résultats ?

- **Regroupement sous un seul numéro :**

Il m'a semblé opportun de garder cette distinction comme base de travail, de ne pas remettre en question la numérotation de Guthrie même si ses critères n'apparaissent pas toujours de façon très claire ou cohérente. Les phonèmes restant stables entre les "starred forms" ont été comptés une seule fois, par contre chaque différence a été prise en compte. À titre illustratif, la Cs 1646 est composée de deux formes à astérisque dont la différence repose sur le timbre de la deuxième voyelle : \*t à d় et \*t à d়. Cette Cs a été comptée deux fois pour tout calcul concernant la voyelle de deuxième syllabe, et une seule fois pour ses autres éléments. Cette démarche est valable pour l'ensemble du bantou commun que la série comparative bénéficie d'une lettre indexée (a, b... ou x, y...) ou non. Ceci entraîne inévitablement des écarts d'un tableau à l'autre, ils sont précisés en note de bas de page.

## 2 - *Cooccurrences étudiées*

Cette analyse possède un caractère exhaustif et systématique lié à la volonté de cerner la structure phonologique des unités lexicales. Chaque occurrence ou cooccurrence traitée est, pour un corpus donné, triée en fonction de la catégorie grammaticale à laquelle appartient l'étyomon : nom vs verbe. Puis, pour chacune de ces deux catégories, l'étude tient compte du schème syllabique auquel appartient la séquence observée.

En préalable, je rappellerai que la structure du mot a été reconstruite sous la forme suivante : [préfixes [racine-suffixes-(voy finale) ]<sub>thème</sub> ] mot

Le mot est composé d'une suite de syllabes ouvertes de type CV ou V. Dans le cas des verbes, la frontière de syllabes ne s'accorde pas forcément avec la frontière de morphème, puisque la voyelle finale n'appartient pas au radical. Le schème canonique des verbes est -CVC-, celui des substantifs est -CVCV.

Le tableau 1 de la page suivante indique comment se combinent les schèmes, et quel type de suites ils engendrent pour chacun des lexiques étudiés. Il est donné en préalable au travail qui va suivre, car chaque analyse s'appuiera sur la distinction nom vs verbe et à l'intérieur de chacune de ces catégories, sur le schème syllabique.

Dans ce tableau, on voit apparaître une grande variété de schèmes, mais la majorité ne renvoie qu'à un nombre très limité de suites qui ne seront observées que dans les cas de distributions manquantes.

	BANTOU COMMUN				PB X		B.L.R.1	
	Cs		ps		noms	verbes	noms	verbes
	noms	verbes	noms	verbes	noms	verbes	noms	verbes
-C-								0,1%
-CV	8,27%	3,27%	5,21%	2,26%	11,67%	5,64%	7,03%	2,97%
-CVV	5,09%		3,99%		5,05%		5,64%	
-CVC-		52,95%		51,41%		63,20%	0,1%	59,55%
-CVVC-		13,42%		20,34%		11,87%		11,14%
-CVCV	75,57%	1,96%	73,6%	0,56%	77,29%	0,89%	77,63%	1,91%
-CVVCV	7,7%	0,57%	7,06%		3,15%		2,34%	0,1%
CVCVV	0,8%		3,68%		0,32%		0,96%	
-CVVCVV	0,08%						0,64%	
-CVCVC-		24,14%		18,64%		18,10%		23,14%
-CVVCVC-		1,63%		2,26%		0,30%		0,74%
-CVCVVC-		0,25%						
-CVVCVVC-		0,08%						
-CVCVCV	2,15%	0,65%	4,9%	1,13%	2,21%		3,4%	
-CVVCVCV			0,61%	0,56%			0,1%	
-CVCVVCV							0,2%	
-CVCVCVC-		1,06%		2,82%				
-CVCVVCVC-								0,1%
-CVCVCVVC	0,32%		0,92%		0,32%		0,42%	
-VC-								0,1%
-VCV-							1,49%	
-VCVC-								0,2%
-VCVCV							0,1%	
	/1254	/1219	/326	/177	/317	/337	/948	/940

Tableau 1

## 2.1 - Cooccurrences analysées

Les trois unités phonologiques (voyelle, consonne, ton) ont été observées en contexte intrasyllabique et intersyllabique. Pour les substantifs, je présenterai la combinatoire des \*-CVCV, des \*-CVV (CV) et des \*-CV. Pour les racines verbales, je pourrai comparer essentiellement les \*-CVC-, \*-CVCVC-, \*-CVVC-.

Les cooccurrences suivantes seront analysées :

- **Les cooccurrences consonnes/voyelles** qui vont quantifier les séquences  $*C_1/*V_1$ ,  $*V_1/*C_2$ ,  $*V_{12}/*C_2$ ,  $*C_2/*V_2$  et  $*V_2/*C_3$ . Ces cooccurrences entre segments adjacents seront résumées en partie 2.
- **Les cooccurrences vocaliques**, soit  $*V_1/*V_2$ ,  $*V_{11}/*V_{12}$  et  $*V_{12}/*V_2$ . Ces cooccurrences participeront à la discussion du chapitre 1 sur l'harmonie vocalique, en partie 3.
- **Les cooccurrences consonantiques**, soit  $*C_1/*C_2$ ,  $*C_2/*C_3$  et  $*C_1/*C_3$ . L'analyse de ces cooccurrences permettra de soulever une nouvelle problématique autour des contraintes consonantiques dans le chapitre 2 de la troisième partie.
- **Les cooccurrences consonnes/tons**, soit  $*C_1/*T_1$  et  $*C_2/*T_2$  et les **Cooccurrences voyelles/tons**  $*V_1/*T_1$  et  $*V_2/*T_2$  seront en partie 3, chapitre 3.

## 2.2 - Bilan

Tout cela nous conduit à un nombre élevé de tableaux, dont le degré de pertinence est inégal. Une partie d'entre eux servira de point d'appui aux discussions de la partie 3, d'autres apparaîtront en partie 2 pour préciser les contraintes relevées dans les reconstructions.

Enfin un dernier groupe de tableaux sera consultable uniquement dans les annexes. L'annexe 1 contient les tableaux extraits des données des Cs et de B.L.R.1. Cette annexe présentera les tableaux d'occurrences ayant permis le calcul des pourcentages, et les tableaux dont l'ordonnancement des segments diffère. Ces tableaux ont donc un numéro identique au tableau auquel ils renvoient dans le corps du texte. Dans plusieurs situations : fréquences totales rares, faible pertinence, les tableaux des Cs et de B.L.R.1 sont exclusivement en annexe. Dans ce cas leur numérotation suit le plan de la thèse et elle est indiquée en note de bas de page.

L'annexe 2 regroupera les fréquences du PB X et les tableaux seront numérotés à partir de 201.

L'annexe 3 rassemblera les tableaux distributionnels des séries partielles avec une numérotation à partir de 301.

### **3 - Index des tableaux**

Cet index précise le numéro attribué à chaque tableau. Les tableaux des Cs et de B.L.R.1 sont dans le corps de la thèse et-ou en annexe. Comme ils sont numérotés en fonction de leur ordre d'apparition, le numéro devrait permettre de les retrouver rapidement.

#### **Occurrences consonantiques : \*C<sub>1</sub>**

- Racines verbales

B.L.R.1	Tableaux n° 2
Cs	Tableaux n° 3
PB X	Tableaux n° 201
Ps	Tableaux n° 301

- Thèmes nominaux

B.L.R.1	Tableaux n° 4
Cs	Tableaux n° 5
PB X	Tableaux n° 202
Ps	Tableaux n° 302

#### **Occurrences consonantiques : \*C<sub>2</sub>**

- Racines verbales

B.L.R.1	Tableaux n° 6
Cs	Tableaux n° 7
PB X	Tableaux n° 203
Ps	Tableaux n° 303

- Thèmes nominaux

B.L.R.1	Tableaux n° 9
Cs	Tableaux n° 8
PB X	Tableaux n° 204
Ps	Tableaux n° 304

#### **Occurrences consonantiques : \*C<sub>3</sub>**

• Racines verbales

B.L.R.1	Tableaux n° 10
Cs	Tableaux n° 11
PB X	Tableaux n° 205
Ps	Tableaux n° 305

• Thèmes nominaux

B.L.R.1	Tableaux n° 12
Cs	Tableaux n° 13
PB X	Tableaux n° 206
Ps	Tableaux n° 306

**Cooccurrences \*C<sub>1</sub>/\*V<sub>1</sub>**

• Racines verbales

B.L.R.1	*-CVC-	Tableaux n° 14
	*-CVCVC-	Tableaux n° 15
	*-CVVC-	Tableaux n° 16
Cs	*-CVC-	Tableaux n° 17
	*-CVCVC-	Tableaux n° 18
	*-CVVC-	Tableaux n° 19
PB X	*-CVC-	Tableaux n° 207
	*-CVCVC-	Tableaux n° 208
	*-CVVC-	Tableaux n° 209
Ps	*-CVC-	Tableaux n° 307
	*-CVCVC-	Tableaux n° 308
	*-CVVC-	Tableaux n° 309

• Thèmes nominaux

B.L.R.1	*-CVCV	Tableaux n° 20
	*-CV	Tableaux n° 21
	*-CVV	Tableaux n° 22
Cs	*-CVCV	Tableaux n° 23
	*-CV	Tableaux n° 24
	*-CVV	Tableaux n° 25
PB X	*-CVCV	Tableaux n° 210
	*-CV	Tableaux n° 211
Ps	*-CVCV	Tableaux n° 310

## Cooccurrences \*C<sub>2</sub>/\*V<sub>2</sub>

- Thèmes nominaux

B.L.R.1	*-CVCV	Tableaux n° 28
Cs	*-CVCV	Tableaux n° 29
PB X	*-CVCV	Tableaux n° 212
Ps	*-CVCV	Tableaux n° 311

## Cooccurrences \*V<sub>1</sub>/\*C<sub>2</sub>

- Racines verbales

B.L.R.1	*-CVC-	Tableaux n° 30
	*-CVCVC-	Tableaux n° 31
Cs	*-CVC-	Tableaux n° 32
	*-CVCVC-	Tableaux n° 33
PB X	*-CVC-	Tableaux n° 213
	*-CVCVC-	Tableaux n° 214
Ps	*-CVC-	Tableaux n° 312
	*-CVCVC-	Tableaux n° 313

- Thèmes nominaux

B.L.R.1	*-CVCV	Tableaux n° 34
Cs	*-CVCV	Tableaux n° 35
PB X	*-CVCV	Tableaux n° 215
Ps	*-CVCV	Tableaux n° 314

## Cooccurrences \*V<sub>2</sub>/\*C<sub>3</sub>

- Racines verbales

B.L.R.1	*-CVCVC-	Tableaux n° 36
	*-CVC-VC-	Tableaux n° 38
Cs	*-CVCVC-	Tableaux n° 37
PB X	*-CVCVC-	Tableaux n° 216
Ps	*-CVCVC-	Tableaux n° 315

## Cooccurrences \*V<sub>12</sub>/\*C<sub>2</sub>

### • Racines verbales

B.L.R. 1	*-CVVC-	Tableaux n° 39 -41
Cs	*-CVVC-	Tableaux n° 40-42
PB X	*-CVVC-	Tableaux n° 217-218
Ps	*-CVVC-	Tableaux n° 316

## Cooccurrences \*Pn/\*C<sub>1</sub>

### • Thèmes nominaux

B.L.R.1	*-CVCV	Tableaux n° 44
Cs	*CVCV	Tableaux n° 45
PB X	*-CVCV	Tableaux n° 219

## Cooccurrences \*V<sub>1</sub>/\*V<sub>2</sub> et \*V<sub>11</sub>/\*V<sub>12</sub>

### • Thèmes nominaux

B.L.R. 1	*-CVCV	Tableaux n° 46
	*-CVV	Tableaux n° 48
	*-CVVCV	Tableaux n° 50
Cs	*-CVCV	Tableaux n° 47
	*-CVV	Tableaux n° 49
	*-CVVCV	Tableaux n° 51
PB X	*-CVCV	Tableaux n° 220
Ps	*-CVCV	Tableaux n° 317

### • Racines verbales

B.L.R. 1	*-CVVC-	Tableaux n° 52
	*-CVCVC-	Tableaux n° 54
Cs	*-CVVC-	Tableaux n° 53
	*-CVCVC-	Tableaux n° 55
PB X	*-CVVC-	Tableaux n° 221
	*-CVCVC-	Tableaux n° 222
Ps	*-CVVC-	Tableaux n° 318
	*-CVCVC-	Tableaux n° 319

## Cooccurrences \*C<sub>1</sub>/\*C<sub>2</sub>

• Racines verbales

B.L.R.1	*-CVC-	Tableaux n° 61
	*-CVCVC-	Tableaux n° 63
	*-CVC-+*-CVCVC-	Tableaux n° 65
	*-CVVC-	Tableaux n° 67
Cs	*-CVC-	Tableaux n° 62
	*-CVCVC-	Tableaux n° 64
	*-CVC-+*-CVCVC-	Tableaux n° 66
	*-CVVC-	Tableaux n° 68
PB X	*-CVC-	Tableaux n° 223
	*-CVCVC-	Tableaux n° 224
	*-CVVC-	Tableaux n° 225
Ps	*-CVC-	Tableaux n° 320
	*-CVCVC-	Tableaux n° 321
	*-CVVC-	Tableaux n° 322

• Thèmes nominaux

B.L.R.1	*-CVCV	Tableaux n° 73
Cs	*-CVCV	Tableaux n° 74
	*-CVVCV	Tableaux n° 75
PB X	*-CVCV	Tableaux n° 228
Ps	*-CVCV	Tableaux n° 323

### Cooccurrences \*C<sub>2</sub>/\*C<sub>3</sub>

• Racines verbales

B.L.R.1	*-CVCVC-	Tableaux n° 69
Cs	*-CVCVC-	Tableaux n° 70
PB X	*-CVCVC-	Tableaux n° 226

• Thèmes nominaux

B.L.R.1		Tableaux n° 78
Cs		Tableaux n° 79

### Cooccurrences \*C<sub>1</sub>/\*C<sub>3</sub>

• Racines verbales

B.L.R.1	*-CVCVC-	Tableaux n° 71
Cs	*-CVCVC-	Tableaux n° 72

PB X \*-CVCVC- Tableaux n° 227

- Thèmes nominaux

B.L.R.1 Tableaux n° 76

Cs Tableaux n° 77

### Cooccurrences \*C<sub>1</sub>/\*T<sub>1</sub>

- Racines verbales Tableaux n° 80

- Thèmes nominaux Tableaux n° 81

### Cooccurrences \*V<sub>1</sub>/\*T<sub>1</sub>

- Racines verbales

Cs \*-CVC- Tableaux n° 82

\*-CVCVC- Tableaux n° 83

\*-CVVC- Tableaux n° 84

PB X \*-CVC- Tableaux n° 229

\*-CVCVC- Tableaux n° 230

\*-CVVC- Tableaux n° 231

Ps \*-CVC- Tableaux n° 324

\*-CVCVC- Tableaux n° 325

- Thèmes nominaux

Cs \*-CVCV Tableaux n° 85

\*-CV Tableaux n° 86

\*-CVV Tableaux n° 87

PB X \*-CVCV Tableaux n° 232

Ps \*-CVCV Tableaux n° 326

### Cooccurrences \*V<sub>2</sub>/\*T<sub>2</sub>

- Thèmes nominaux

Cs \*-CVCV Tableaux n° 88

PB X \*-CVCV Tableaux n° 233

Ps \*-CVCV Tableaux n° 327

## Cooccurrences \*C<sub>1</sub>/\*V<sub>1</sub>/\*T<sub>1</sub>

- Racines verbales

Cs	*-CVC-	Tableaux n° 89
	*-CVCVC-	Tableaux n° 90
PB X	*-CVC-	Tableaux n° 234
	*-CVCVC-	Tableaux n° 235

- Thèmes nominaux 91