

Bible & Data Mining

[Analyse, Stylométrie]

Kharchaf Idris

Professeur de l'enseignement supérieur en Analyse des données
Département de Mathématiques, Faculté des sciences, Université M^d V, Rabat

Présentation

Le texte de la Bible hébraïque se présente comme une collection de livre d'époques et d'auteurs très divers. Il a été transmis oralement durant une période (au moins deux siècles).

Les églises chrétiennes classent les livres de la Bible en deux grandes parties :

✚ L'ancien Testament.

✚ Le nouveau Testament.

La recherche dont s'inspire le présent travail a pour objectif principal, d'analyser l'ensemble des livres de la Bible, et de connaître la relation qui existe entre les différents livres de la Bible, en faisant recours aux techniques de Data Mining (construction d'un arbre hiérarchique).

Le recours aux techniques de l'intelligence artificielle (Analyse des données bibliques), comme un moyen de gérer des données textuelles, nous permet de mettre en évidence des faits patents dans la masse des données à traiter.

Ici, nous avons analysé un tableau de correspondance entre un ensemble de 508 mots et un ensemble de 73 livres de la Bible.

Définition :

La description de la Bible peut se mettre sous forme d'un tableau rectangulaire de nombres positifs, l'élément $K(m_i, l_j)$ situé à l'intersection de la ligne m_i (mots) et de la colonne l_j (livres), représente la fréquence du mot m dans le livre l .

Livres Mots	l_1	l_2	$l_3 \dots$	$\dots l_j \dots$	l_{72}	l_{73}
m_1						
m_2						
m_3						
:						
m_i				$k(m_i, l_j)$		
:						
:						
m_{508}						

(Fig.1)

Donc les k_{m_l} sont tous des quantités homogènes, exprimées avec la même unité de mesure, autrement dit, $k_{m_l} =$ nombre de fois que le mot m se rencontre dans le livre l .

Méthodologie

Pour échapper aux contraintes des méthodes d'analyses statistiques dites paramétriques, au modèle de référence imposé par avance, nous avons tenté d'utiliser, pour résoudre les problèmes textuels posés par l'unité ou la pluralité supposée d'auteur dans un texte donné, autant que pour la classification typologique des données textuelles, des méthodes d'analyse mises au point par le professeur J.P. Benzecri (*Directeur du Laboratoire de statistique et de l'analyse des données, Université Pierre & Marie curie, Paris VI, France*), bien qu'il existait d'autres chercheurs comme Sokal & Sneath, principales of numerical Taxonomy, 359p, Freeman & Co 1963.

Les tableaux de contingence qui doivent nécessairement être établis pour être soumis à ces méthodes d'analyse, permettent, dans le cadre de l'analyse de la distribution des 508 mots les plus fréquents d'un livre biblique donné, de conserver la division capitulaire à laquelle est habitué tout lecteur de la Bible. 4

La représentation graphique des résultats, est réalisée sous la forme de projections dans un plan délimité par deux axes, choisis parmi les axes factoriels portant l'information calculée par le programme. Cette représentation qui résume visuellement l'information, à laquelle s'ajoutent les tableaux édités par le programme d'analyse automatique que l'on doit consulter pour interpréter correctement la représentation graphique.

Plan de l'étude

Nous avons choisi de mener notre étude en trois étapes :

Etape1(E1) -Une étude globale, portant sur les 73 livres de la Bible, nous permettra de situer les mots les plus employés (en général) et les livres les uns par rapport aux autres ; Nous disposons donc d'un tableau (Matrice) de 508 lignes et de 73 livres.

Etape2(E2) -Une construction hiérarchique basée sur le résultat de l'analyse factorielle des correspondances, autrement dit la construction d'un arbre hiérarchique d'un tableau (Matrice) de 508 lignes et 7 facteurs.

Etape3(E3) - La stylométrie

E1 : Les Données de Base (D.B.)

Les résultats de l'analyse des correspondances des données bibliques, donnent une information dont la nature est proche de celle fournie par les techniques d'*Analyse Factorielle*.

Partons donc d'une matrice de 508 lignes et de 73 livres, est une table de fréquences à double entrée, c'est-à-dire une matrice de contingence sous la forme $K(508,73)$ (Fig.1).

Partons du tableau ($M * L$), croisant avec l'ensemble de mots M (508 mots) (Tab.1), l'ensemble des livres de la Bible (73 livres) (Tab.2) : $k(m, \ell)$ = nombre d'occurrences du mot « m » dans le livre « ℓ ».

E1 :1. Les Mots (M) utilisés de la Bible et les abréviations qui les désignent

L'étude qui a été faite est basée sur la table pastorale*, qui constitue une présentation méthodique des mots (en langue française), des idées, des sujets et des noms principaux du livre saint (ces derniers ont été pris selon leur ordre alphabétique, commençant par Aaron et terminant par Zakaria).

La table pastorale sert à retrouver dans la Bible tel ou tel passage à partir des mots principaux qu'il contient.

Donc les textes contenus dans la table pastorale sont de deux sortes :

- Les citations verbales, où se retrouve le mot-clef.

- Les citations idéologiques qui en contiennent l'idée.

Mot	Abréviation	Mot	Abréviation	Mot	Abréviation
aaron	aaa	Abaissement	a1	abandon	a2
abba	a3	Abdenago	aab	abdon	Aac
abeille	aad	Abel	aae	Abia	Aaf
abiam	aag	Abias	aah	abiatar	Aai
abiezer	aaj	abigail	aak	abinadab	Aal
abiram	aam	Abisag	aan	abisai	Aao
abiu	aap	Abiud	aaq	ablution	Aar
aabner	aas	Abminable	aat	abraham	Aau
abri	aav	Absalom	aaw	Acan	Aax
accepter	aay	Acclamation	aaz	accomplir	Aca
accueil	acb	Accusation	acc	achab	Acd
achaie	ace	Achat	act	achaz	Acg
Acheter	ach	Achim	aci	achimaas	Acj
achimelec	ack	Achinoam	acl	achior	Acm
Achis	acn	Achipothel	aco	Acor	Acp
acquérir	acq	Acsa	acr	Acte	Acs
Action	act	Benediction	acu	ben hinnon	Acv
Dieu beni un peuple	acw	la creation deni dieu	acx	beni soit le seigneur	Acy
l'homme benit Dieu	acz	Benjamin	ada	Beor	Adb
bera	adb	Beree	add	berzellai	Ade
Dieu beni israel	adf	Betail	adg	bethanie	Adh
bethaven	agi	Bethel	agj	bethleem	Agk
bethoron	agl	Bethsabee	agm	nethsaid	Agn
bethulie	ago	betosomes	agp	betsur	Agq
betzakar	agr	Bien	ags	biefaisant	Agt
bien heureux	agu	Bierre	agv	bijou	Agw
bitume	agx	Blanc	agy	blacheme	Agz
bœuf	aha	boire du vin	ahb	adam	Ahc
adar	ahd	Administrati on	ahc	Adolescen ce	Ahf
adonat	ahg	Adonias	ahh	adoration de dieu	Ahi
adoration de jesus	ahj	Adversaire	ahk	Age	Ahl
agenouiller	ahm	Agneau	ahn	agreable a dieu	Aho

ahia	ahp		Ahicam	ahq		ahilud	Ahr
.....
....
zorobabel	azf		Zara	azi		zacharie	Azj

(Tab .1) (*) L'orthographe des noms propres est généralement celle adoptée par la Bible de Maredsus (Edition,1968),
G. Passelecq & F. Poswick (Table pastorale de la Bible), Ed .P. Lethielleux 1974 ;1214 pages.

E1:2. Liste des variables utilisées

Nom	N°	Abréviation	Nombre de chapitres	Nature
Genèse	1	GN	50	Pentateuque (A.T)
EXODE	2	EX	40	:
LEVITIQUE	3	LV	27	:
NOMBRES	4	NB	36	:
DETEURONOMIE	5	DT	34	:
L .DE JOSUE	6	JOS	24	Livre historique (A.T)
L. JUGES	7	JUG	21	::
L.RUTH	8	RUTH	4	:
1°L. SAMUEL	9	1 SA	31	::
2°L. « « «	10	2 SA	24	:
1° L. ROIS	11	1 RO	22	:
2° L. ROIS	12	2 RO	25	::
1° L. de chronique	13	1 CH	29	::
2° L. « « «	14	2 CH	36	:
L. D'ESDRAS	15	ESD	10	;
L. de NEHEMIE	16	NEH	13	:
L. de ESTHER	19	EST	10	:
L. de JOB	22	JOB	42	:
L. de PSAUMES	23	PS	150	Livre poétique (A.T)
Les proverbes	24	PR	31	:
Qohèleth ou livre d'ECCLESIASTE	25	QO	12	:
Cantique des cantiques	26	CT	8	:
ISAIE	29	IS	66	Prophète (A.T.)
JEREMIE	30	JER	52	:
Lamentation	31	LAM	5	:
EZECHIEL	33	EZ	48	:

DANIEL	34	DAN	12	:
OSEE	35	OS	14	:
JOEL	36	JOEL	3	:
AMOS	37	AM	9	:
ABDIAS	38	1BD	1	:
JONAS	39	JON	4	:
MICHEE	40	MI	7	:
NAHUM	41	NAH	3	:
HABAKUK	42	HAB	3	:
SOPHONIE	43	SO	3	::
AGGEE	44	AG	3	:
ZACKARIE	45	ZAC	14	:
MALACHIE	46	MAL	4	:
L.de TOBIE	17	TOB	14	Deutérocannonique (A.T)
L. de JUDITH	18	JUD	16	« « « ou Apocryphes
1°L. de Maccabées	20	1 MA	16	:
2° L. « « «	21	2 MA	15	:
L. de la SAGESSE	27	SG	19	:
Le SIRACIDE	28	SIR	51	:
BARUCH	32	BAR	5	:
EVANGILE selon MATHIEU	47	MATT	28	Nouveau testament
EVANGILE selon MARC	48	MARC	16
E. LUC	49	LUC	24
E. JEAN	50	JEAN	21	:
Actes des apôtres	51	ACT	28	:
Epître aux romains	52	ROM	16	:
1°Epître aux corinthiens	53	1 CO	16	:
2° « « «	54	2 CO	13	::
Epître aux galates	55	GAL	6	:
Epître aux Ephésiens	56	EPH	6	:
Epître aux philippiens	57	PHI	4	:
Epître aux Colossiens	58	COL	4	:
1° Epître aux Thessaloniens	59	1 TH	5	:
				:

2° Epître aux Thessaloniens	60	2 TH	3	:
1° Epître à Timothée	61	1 TI	6	:
2 Epître à Timothée	62	2 TI	4	:
Epître à Tite	63	TITE	3	:
Epître à Philémon	64	PHI	1	:
Epître aux hébreux	65	HEB	13	:
Epître de Jacques	66	JAC	5	:
1° Epître de Pierre	67	1 PI	5	:
2° Epître de Pierre	68	2 PI	3	:
1° Epître de Jean	69	1 JN	5	:
2° Epître de Jean	70	2 JN	1	:
3° Epître de Jean	71	3 JN	1	:
Epître de Jude	72	JUDE	1	:
Apocalypse de Jean	73	JEAN	22	::

(Tab.2)

E1 :3. Résultat de l'Analyse

Un premier travail portant sur la totalité des 73 livres, a fait ressortir avec une très forte contribution quelques mots cités dans les livres de la Bible.

Donc, nous obtenons une sorte de statistique globale grâce aux valeurs propres (λ_α) qui représentent les contributions absolues des mots m à l'axe α , autrement dit : $\lambda_\alpha = \sum \{f_m \cdot F_\alpha(m)^2 / m \in M\}$ (Tab.3).

$\tau_\alpha = \lambda_\alpha / \text{Inert}_{f(s)} \mathcal{N}(M)$; $\text{Inert}_{f(s)} \mathcal{N}(M) = \sum \{\lambda_s / s=1,7\}$ (i.e. la trace par rapport aux 7 facteurs calculés = $\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_7$).

τ_α mesure la part afférente à l'axe α dans l'inertie totale du nuage.

Aussi la première valeur propre « triviale » qui est calculée pour vérification $\lambda_0 = 0.9998909$ (Tab.3)

Nous avons une autre sortie, tableau qui donne plusieurs informations comme la contribution relative (CTR), inertie de chaque livre (INR) et pour chaque mot, le poids, etc. ces paramètres servent à contrôler les résultats de l'analyse.

E1 : 4. Nuages $\mathcal{N}(M)$ & $\mathcal{N}(\mathcal{L})$

E1 :4.1. Nuages des mots $\mathcal{N}(M)$ (Plan : 1*2)

On constate la concentration du nuage des mots au voisinage de l'origine des axes factoriels (fig.2), sauf l'ensemble des mots $M = \{\text{Amanihman, Esther, Adar}\}$ cela est dû à la forte corrélation du facteur F_1 aux mots de l'ensemble M (Tab.4).

Premier facteur : F_1 # Facteur d'ESTHER ($\lambda_1=074$; $\tau_1= 7.413\%$)

En général, le premier axe factoriel se montre comme un levier avec deux bras nettement inégaux (fig.2), l'un (l'axe négatif) est plus étendu vers l'extrémité du plan, alors que l'autre (l'axe positif) est plus condensé autour du centre des axes factoriels.

En bref, nous pouvons dire que le premier axe (premier facteur F_1) est déterminé par l'ensemble $C = \{\text{Adar, Amanihman, Esther}\}$, dont la corrélation du premier facteur (F_1) aux mots de C est très forte, autrement dit : $0.978 \leq \text{corr}[(F_1, c_i) / i=1,3] \leq 1$ (Tab.4) & (fig.2,2 bis), par contre, la contribution relative du facteur F_1 aux reste des mots $D = \{\text{Simon, Judas, Démétrius, Jonathan, Alexandre, Betour, Eléphant, Antiochos, Betzakar, Bacchidés}\}$ est très faible.

Mot	Inertie(λ_m) = $\sum\{F_m^2 \cdot f_m / m \in M\}$	Corrélation(COR)	Contribution relative (CTR)
Adar	3	978	46
Amanihman	14	1000	190
Esther	56	1000	753

(Tab.4)

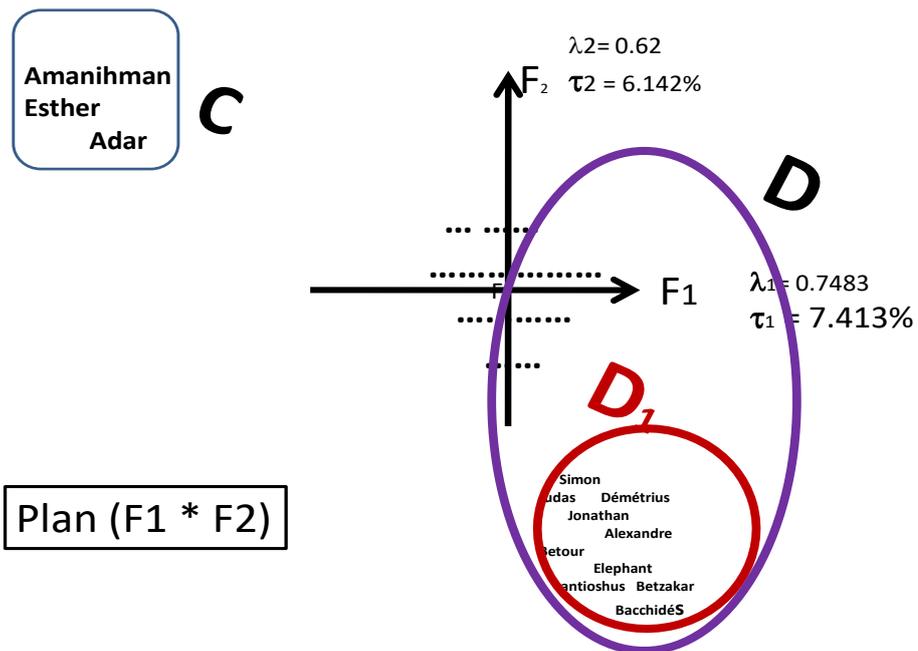
Iner(λ_m) : (Inertie du mot(m)) = (λ_m) = $\sum_m \{F_m^2 \cdot f_m\}$ (**Tab.4**)

COR: représente la corrélation du facteur α au mot m :

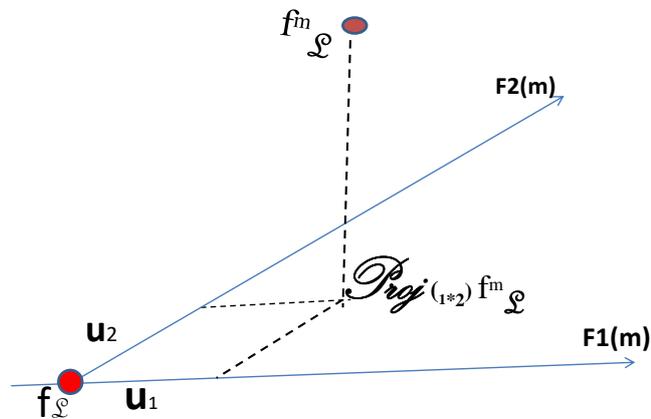
$$= F_{(m)}^2 / \rho_{(m)}^2 = F_{(m)}^2 / d^2(f_{\mathcal{L}}, f_{\mathcal{L}}^m)$$

CTR : est la contribution relative du mot m à l'axe (α) ;

$$\text{Ctr}(m,\alpha)=f_m \cdot F_\alpha(m)^2/\lambda_\alpha ,$$



Nuage des mots $\mathcal{N}(M)$ (fig.2)



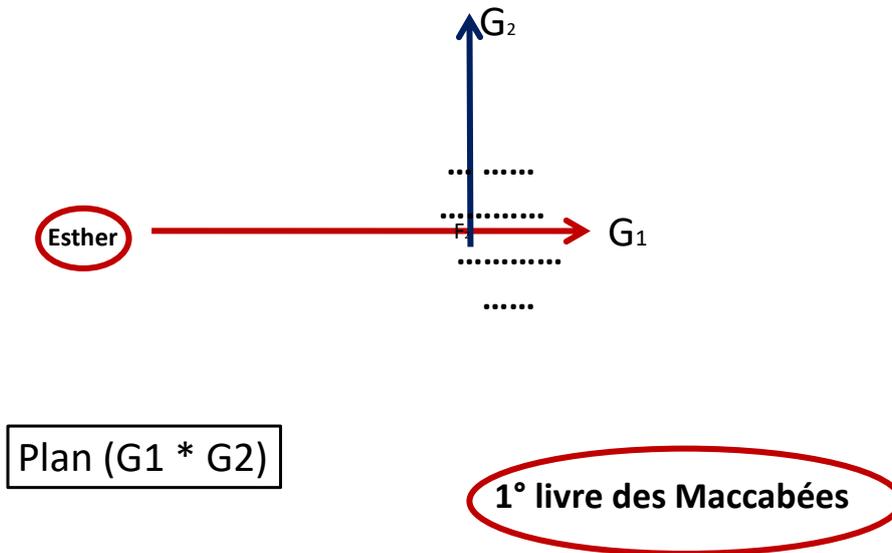
Représentation géométrique d'un point quelconque de m sur un espace à deux dimensions $F_1(m) * F_2(m)$ (fig.2 bis)

La contribution relative du reste des mots $D_1 = \{d_{j/j=1,10}\} \subset D$ est très faible : $0 \leq \text{Cor}(F_1, D_1) < \beta < \text{cor}(F_1, C = \{c_{i=1,3}\}) \leq 1$; $\beta \in \mathbb{R}^+$

Dans le plan de nuage des mots $\mathcal{N}(M)$ (fig.2), apparaissent deux blocs bien distincts, qui ont de fortes contributions : $C = \{\text{Amanihman, Esther, Adar}\}$, et $D = \{\text{Simon, Judas, Démétrius, Jonathan, Alexandre, Betour, Elephant, antioshus, Betzakar, Bacchidés}\}$

E1 :4.2. Nuages des livres $\mathcal{N}(L)$ (fig.3)

Dans le nuage des livres, on trouve la même distribution précédente c'est-à-dire, on constate l'isolement du livre d'Esther (Livre historique), ce livre prophétique qui constitue un élément de l'ancien testament.



Nuage des livres $\mathcal{N}(L)$ (fig.3)

Interprétation

Ce livre est parmi les livres dits historique (Deutéronomique), la relation qui existe entre les mots: Adar, Esther, Amanihman, montre bien l'histoire d'Esther. L'auteur qui a écrit ce livre, nous raconte la deuxième année du règne d'Artaxerxés (fig.4).

Livre d'ESTHER

Amanihman		0 Valeurs faibles ou nulles
Adar		
Esther	Valeurs faibles ou nulles 0	

(Fig.4)

Il y avait dans la ville de sous-la-citadelle un juif nommé Mardochée descendant de Yaïr, étant tuteur d'un enfant, une fille de son oncle Amina Dad, elle se nommait Esther (avait perdu ses parents).

Le livre d'Esther insiste sur un point essentiel, c'est Dieu qui agit et qui sauve, il est le seul, Sauveur de l'être humain lorsqu'il est livré aux mains de ses ennemis. Pour cela, l'épopée des Maccabées choisit les moyens les plus faibles ; la main d'une femme (fig.3).

Concrètement, le nom d'Esther apparaît avec une très haute fréquence dans un seul livre, on a donc un tableau doublement partitionné. On peut en déduire que le livre d'ESTHER est fortement aramaisé et plaignant un style deutéronomique archaïque.

En plus, ce livre d'ESTHER eut quelque peine à pénétrer dans le canon juif.

Deuxième facteur (fig.2): F₂ # Facteur des Maccabées

$$(\lambda_2=0.62 ; \tau_2= 6.1412\%)$$

A/ Par rapport au nuage des mots $\mathcal{N}(\mathbf{M})_{F_2}$

Pour le deuxième facteur, on constate la formation de deux blocs :

A .1. Partie négative ($F_2 < 0$) (en bas de la figure 2)

Sur ce facteur négatif : $F_2 < 0$, (détachés du bloc C), on constate le rôle important de l'ensemble des mots D_1 qui engendrent le facteur F_2 , où D_1 : {Simon, Judas, Démétrius, Jonathan, Alexandre, Betsur, Elephant, Antiochos, Betzakar, Bacchidés } (Tab.5).

Mot	COR	CTR		Mot	COR	CTR
Jonathan	954	278		Démétrius	949	121
Bacchidés	954	86		Simon	923	26
Batzacar	954	13		Betsur	935	10
Antiochos	952	120		Judas	777	109
Alexandre	952	95		Elephant	675	5

(Tab.5)

Les contributions des six mots : {Jonathan Bacchidés, Antiochos, Batzacar, Alexandre, Démétrius, Judas} reflètent l'importance de ces mots dans le facteur 2(Tab.5).

A. 2. Partie positive $F_2 > 0$ (en haut de la figure 2)

Sur ce facteur on remarque que, les mots cités sur la partie positive de l'axe 2(en haut de la figure 2), $C = \{\text{Amanihman, Esther, Adar}\}$, ne jouent aucun rôle dans $F_2 > 0$.

B. Par rapport aux nuages des livres $\mathcal{N}(\mathcal{L})_{F_2}$

On constate d'après le calcul fourni par la machine, que le livre le mieux représenté est la variable ℓ_{20} (Tab.6), représente le 1^o livre des Maccabées (fig.3).

L'auteur qui a écrit le 1^{er} livre des Maccabées vers les années 100, raconte l'histoire des trois prophètes premiers : (Judas :3-9), (Jonathan :9-12), (Simon :13-16).

Variabes (livres)	Poids ($f_m = k_m / K$)	Inertie	Cor	Ctr
1° livre des Maccabées	31	58	581	924

$$f_m = k_m / K ; k_m = \sum \{k(m,\ell) / \ell \in \mathcal{L}\} \text{ (Tab.6)}$$

La rédaction définitive de ces prophètes premiers, a été faite par des scribes marqués par le courant de pensée.

L'auteur, s'est efforcé d'en tirer, à l'usage des gens cultivés, un récit agréable à lire. Il a dû pour cela prendre à l'égard de sa source quelques libertés.

Interprétation

Les livres prophétiques ne sont pas des livres d'histoire, ils ne cherchent pas à reconstituer les événements, mais ils nous indiquent le bon chemin, et que le DIEU, reste fidèle à sa promesse de donner l'aide à condition que le peuple revienne à Lui.

Comme le précédent, ce livre se distingue car il raconte un épisode très particulier de l'histoire juive non repris (ou peu cité) par les autres livres.

Troisième facteur (fig.5) : F₃ # Facteur de l'ancien testament

$$\lambda_3 = 0.60 ; \tau_3 = 5.538\%$$

Sur ce facteur, on constate la séparation nette entre l'ancien testament et le nouveau testament, avec l'isolement du 1° livre des Maccabées sur le facteur G₂ (fig.3) et (fig.6).

Sur le graphique ci-dessous (fig.5), on constate la formation de deux blocs sur l'axe F₃ :

A/ Par rapport aux mots $\mathcal{N}(M)_{F_3}$ (Fig.5)

Sur cet axe (facteur) on constate la constitution de deux blocs : F₃ > 0 (partie droite), et F₃ < 0 (partie gauche) (Tab.7)

A : 1. les mots cités sur la partie positive de l'axe 3 ($F_3 > 0$, en haut de la figure 5) (voir Tab.7).

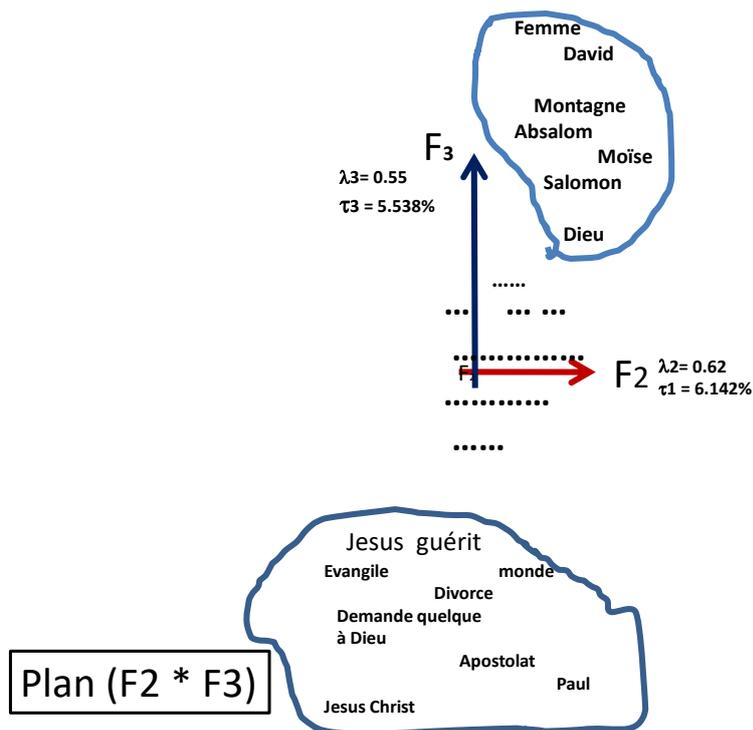
Mots	Corrélation	Contribution
Femme	59	47
David	112	36
Montagne	40	16
Absalom	62	13
Moïse	47	11
Salomon	62	10
Dieu	31	6

(Tab.7)

A :2. les mots cités sur la partie négative de l'axe 3 ($F_3 < 0$, en bas de la figure5) (voir Tab.8).

Mots	Cor	Ctrl
Jésus Christ	657	203
Demande quelque chose à DIEU	118	65
Paul	68	29
Foi	496	27
Vie	285	25
Evangile	378	14
Divorce	48	13
Apostolat	272	11
Jésus guérit	226	10
Monde	282	6

(Tab.8)



Projection de nuage des mots sur le plan (2*3) (fig.5)

B/ Par rapport aux livres $\mathcal{N}_1(L)$

Ici aussi, on trouve une situation identique, à savoir la formation de deux blocs :

B : 1. les livres engendrés par les mots, sur la partie positive de l'axe 3 (partie inférieure de l'axe 3) : $G3 > 0$ (voir Tab.9).

Livres $G3 > 0$ (A.T.)	Cor	Ctr
2° SA	91	64
1° ROI	57	30
CO	50	41
1° SA	69	39
GENESE	44	31
EXODE	49	28
NOMBRES	44	20

Projection des livres sur l'axe 3 négatif (Tab 9)

B : 2. les livres engendrés par les mots sur la partie positive de l'axe (partie inférieure de l'axe 3) : $G3 < 0$ (voir Tab.10).

Livres $G3 < 0$ (A.T.)	COR	CTR
JEAN	349	79
1°PI	145	67
1°JN	254	52
ACT	69	50
MATT	296	49
LUC	109	43
1 TH	177	42
2 CO	199	28
JAC	108	27
1 CO	255	22
MARC	140	22
COL	117	20
ROM	224	23
EPH	196	19

Projection des livres sur l'axe 3 négatif (Tab.10)

Les tableaux (Tab .9, 10) indiquent uniquement les livres qui ont une forte corrélation avec le 3° axe (Troisième facteur)

Interprétation

Sur ce facteur (Axe 3), on remarque que l'ancien testament regroupe des livres très divers, qui nous racontent de David, Absalon, et de la jeune femme moabite qui est devenue la bisaïeule du roi David, tout cela comme témoin privilégié des relations de Dieu, et ceux qu'il appelle à entrer sans son grand dessein.

Sur la partie négative de l'axe 3 ($G3 < 0$), le nouveau testament raconte l'histoire et la vie de Jésus, les évangiles mettent les fidèles en contact direct avec Jésus.

En général, l'évangile s'attache à présenter Jésus comme le sauveur de tous les hommes, son but est d'appeler tous ses lecteurs à croire en

Jésus et à découvrir ainsi la vraie, on peut dire aussi que Jésus est venu pour établir les liens quelque peu distendus en DIEU et son peuple (toujours selon l'Évangile).

Etude comparative

- ✚ Le sens théologique de l'ancien, n'apparaît en définitive qu'à la lumière du nouveau.
- ✚ Le nouveau testament permettra de définir les sens de l'ancien puisque, dans sa nouveauté, il sera un accomplissement des écritures.
- ✚ Les grandes périodes littéraires de l'A.T., elles coïncident dans une mesure avec celle de l'histoire des Israéliens.
- ✚ Le style de N.T. est généralement coulant et limpide ; avec quelques passages plus rocailleux. Assez souvent, comme dans la langue parlée, la phrase brave les règles de la grammaire et de la syntaxe.
- ✚ Les livres de L'A.T. s'enchaînent les uns aux autres et l'on peut y retrouver une trame continue d'événements depuis la création jusqu'à la mort de Moïse.
- ✚ Dans le Nouveau testament, le dialogue est bref, destiné à mettre en relief un mot du sauveur (l'événement est une controverse).
- ✚ Le progrès de la révélation divine lié à l'éducation spirituelle des Israéliens.

Ce sont les raisons pour lesquelles nous avons une séparation nette sur la (figure 5).

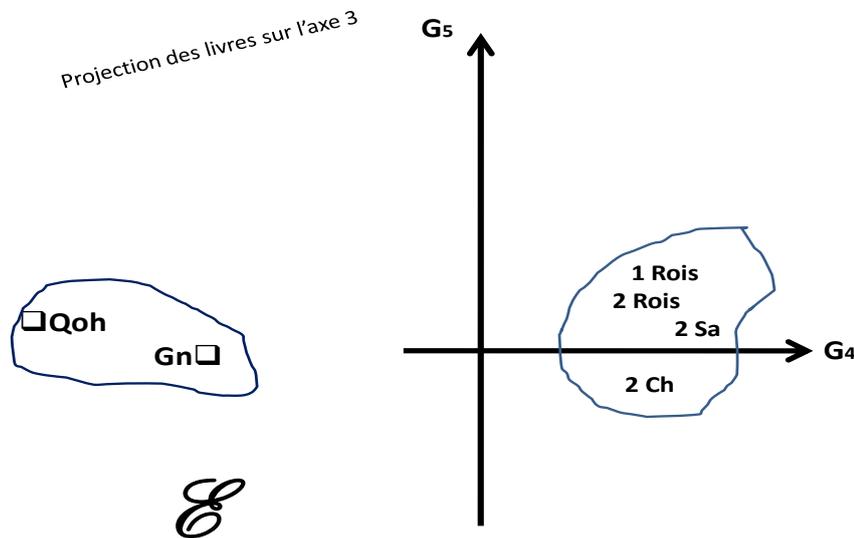
Quatrième facteur (fig.6) : F₄ # Facteur des livres historiques

$$\lambda_4=0.50 ; \tau_2= 4.971\%$$

1/ Par rapport au nuage des livres : $\mathcal{N}(\mathcal{L})_{F4}$

Nous remarquons sur cet axe la séparation des livres de l'A.T. D'une part, le livre de Qohelet (partie négative de l'axe 4) qui est en forte corrélation avec le quatrième facteur, et les livres représentés par l'ensemble : $\mathcal{E} = \{2 SA, 1 Ro, 2 Ro, 2 CH\}$ (Fig.6a).

Le livre de Qohelet (nom hébreu) ou Ecclésiaste est considéré comme un livre étrange, il présente une réflexion non conformiste sur les sens de la vie humaine.



En ce qui concerne l'ensemble des livres sont :

1. A/ sur la partie positive (Livres : \mathcal{L}) (Tab.11;fig.6a)

Variables (Livres) $G_4 > 0$	Cor	Ctr
1°LIVRE DE ROIS	350	291
2°LIVRE SAMUEL	84	62
2°LIVRE DE ROIS	79	48
2° LIVRE DE CHRONIQUES	96	33

(Tab.11) Les variables (livres) et leurs emplacements par rapport à l'axe G_4 positif ($G_4 > 0$)

1. B/Sur la partie négative (Livres : \mathcal{L}) (Tab.12;fig.6a)

Variables (Livres) $G_4 < 0$	Cor	Ctr
QOHELET	470	435
GEN7SE	100	78

(Tab.12) les variables et leurs emplacements par rapport à l'axe G_4 négatif ($G_4 < 0$), (Fig.6a).

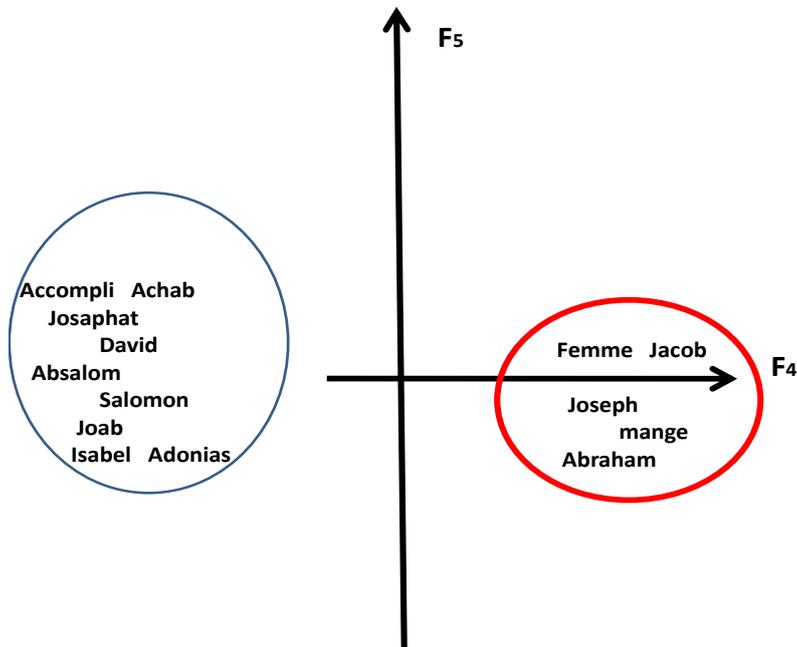
2/ Par rapport au nuage des mots : $\mathcal{N}(\mathcal{M})_{F_4}$

2.A/ Les Mots et leurs emplacements par rapport à l'axe $F_4 > 0$.

(Tab.13 ; Fig.6b).

Individus(mots) $F_4 > 0$	Iner	Cor	Ctr
Femme	44	483	426
Jacob	8	85	14
Abraham	9	82	14
Joseph	7	86	12
Mange	2	348	13

Tableau de Contributions et corrélations des mots sur quatrième axe positif (Tab.13)



Projection de \mathcal{N} sur les Plan (4*5) (fig.6b)

2.B les mots et leurs emplacements para rapport à l'axe F_4 négatif($F_4 < 0$) (Tab. 14 ; fig.6b).

Individus(mots) $F_4 < 0$	Iner	Cor	Ctr
Accompli	10	321	63
Salomon	3	342	61
Achab	5	370	37
David	18	99	35
Elie	4	371	27
Hiram	3	348	20
Absalom	11	65	15
Joab	6	106	14
Josaphat	5	154	16
Adonias	2	302	10
Jezabel	1	326	9

Tableau d'inertie de Contributions et corrélations des mots sur quatrième axe négatif (Tab.14)

La correspondance qui existe entre les mots et les livres, qui ont de très fortes contributions avec l'axe 4 est expliquée par les chapitres suivants : $\mathcal{F} = \{\text{ch12}, \text{ch17}, \text{ch1}(\text{v}_{27}), \text{ch2}(\text{v}_{23}), \text{ch13}(\text{v}_{39}), \text{ch4}(\text{v}_{25}), \text{ch11}(\text{v}_1)\}$

Explication

De la correspondance précédente, nous constatons que les livres des rois et les chroniques, ils analysent le règne de Salomon, sa puissance et sa richesse. Et que Qohelet ou la Genèse veulent présenter une vision globale des relations de DIEU avec son peuple, et donc ils emploient d'autres mots.

L'exemple qui apparaît le plus probant est certainement celui du livre des Rois, où 1 Rois (17-22) et 2 Rois (2-9,10) sont bien marqués par un suremploi des verbes.

A l'analyse on note que 1 Rois (17-22) représente le cycle d'Elie séparé par un chapitre de transition, 2 Rois, 1 du cycle d'Elysée 2 Rois (2-9,10).

Cinquième facteur (fig.) : $\lambda_5=0.47$; $\tau_5= 4.674\%$

F_5 # Facteur de séparation : les livres de Torah et les livres historiques.

Cet axe est caractérisé par la formation de deux blocs : d'une part l'ensemble $E=\{\text{Genèse}, \text{Exode}, \text{Nombres}\}$, dont :

Livres	Inertie	Cor	Ctr
Exode	32	157	107
Genèse	39	63	52
Nombres	25	55	30

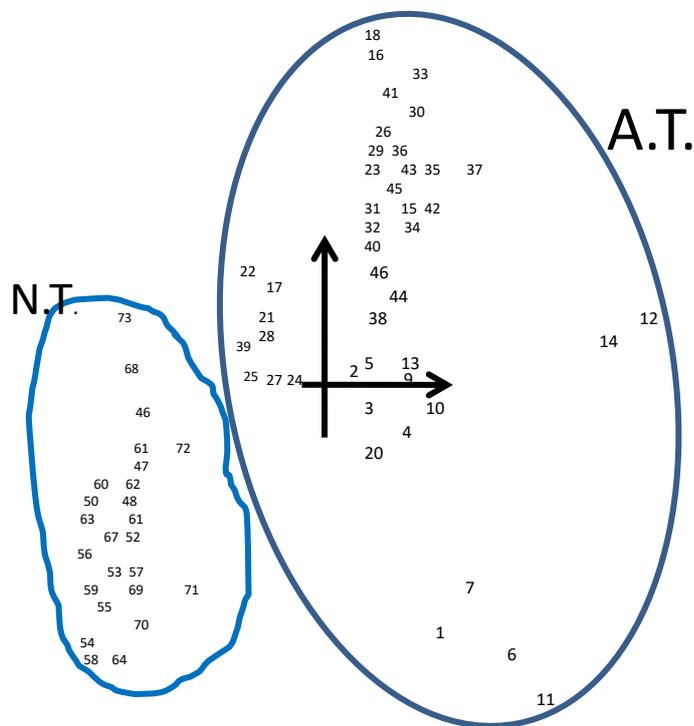
(Tab.15)

On peut remarquer la forte contribution du livre 2(Exode) avec l'axe 5(Tab.15), et l'ensemble $K = \{2Sa, 1Ro, 2Ro, Qo\}$ d'autre part.(Tab.16).

Livres	Inertie	Cor	Ctr
Qohelet	46	386	381
10 Rois	41	127	99
2° Samuel	36	106	71
2°Rois	30	82	73

(Tab.16)

Donc on remarque sur ce facteur, la séparation entre les livres de la loi (Torah) et livres historiques(fig.7).



(Fig.7)

Classification Ascendante hiérarchique

La classification ascendante hiérarchique, est un ensemble de parties d'un ensemble quelconque, partiellement ordonné par inclusion.

La construction de C.A.H. des livres se fait par agrégation successives à partir des éléments de \mathcal{L} .

1/ on réunit les deux éléments les plus proches au sens de la distance (δ) sur l'ensemble \mathcal{L} .

On définit comme une distance distributionnelle entre ℓ et ℓ' , la quantité:
$$\|f_M - f'_M\|_{f_M}^2 = \sum \{ (f_m - f'_m)^2 / f_m \mid m \in M \}$$

2/ On crée ainsi le premier nœud que l'on appelle $\text{Card}\mathcal{L} + 1$.

Soit ℓ l'aîné de ce nœud et ℓ' le benjamin (ℓ et ℓ' étant deux classes terminales).

Donc $\{\ell \text{ et } \ell'\}$ crée sur \mathcal{L} , une partition \mathcal{P} , On calcule une distance δ sur cette nouvelle partition \mathcal{P} .

3/ On calcule ainsi la distance entre la classe $\{\ell \text{ et } \ell'\}$ et les autres classes que $\{\ell\}$ et $\{\ell'\}$.

On crée ainsi une classe sur \mathcal{P} et un deuxième nœud au niveau $n = \text{Card}(\mathcal{L} + 2)$

:
:

i/ On continue jusqu'à ce qu'il ne reste plus qu'une partition de deux classes, telles que la réunion soit l'ensemble \mathcal{L} (entier), $\mathcal{P}_1 = \{ p_i / p_i \subset \mathcal{L} ; i=1,2,3,\dots,n \}$, et on réunit deux parties p_α et p_β de \mathcal{L} qui maximise le moment centré d'ordre 2, d'une partition \mathcal{P}_2 de \mathcal{L} telle que : $\mathcal{P}_2 = \{ p_i / p_i \subset \mathcal{L} ; i=1,2,3,\dots,n \} \cup \{ p_\alpha \cup p_\beta \} - \{ p_\alpha \} - \{ p_\beta \} = \mathcal{P}_1 \cup \{ p_\alpha \cup p_\beta \} - \{ p_\alpha \} - \{ p_\beta \}$.

Application

La méthode de la classification ascendante hiérarchique forme le deuxième volet de l'analyse des données.

Le tableau sur lequel fut effectué cette classification ascendante hiérarchique est le suivant : (70 livres * 7 facteurs) ; (Nous passons par l'intermédiaire des 7 facteurs issus de l'analyse factorielle des correspondances).

Avant d'effectuer la classification des livres, nous avons enlevés tous les profils qui ont des poids nuls, et par conséquent des contributions nulles, sur les axes factoriels, comme le cas du livre 3° Epître de Jean.

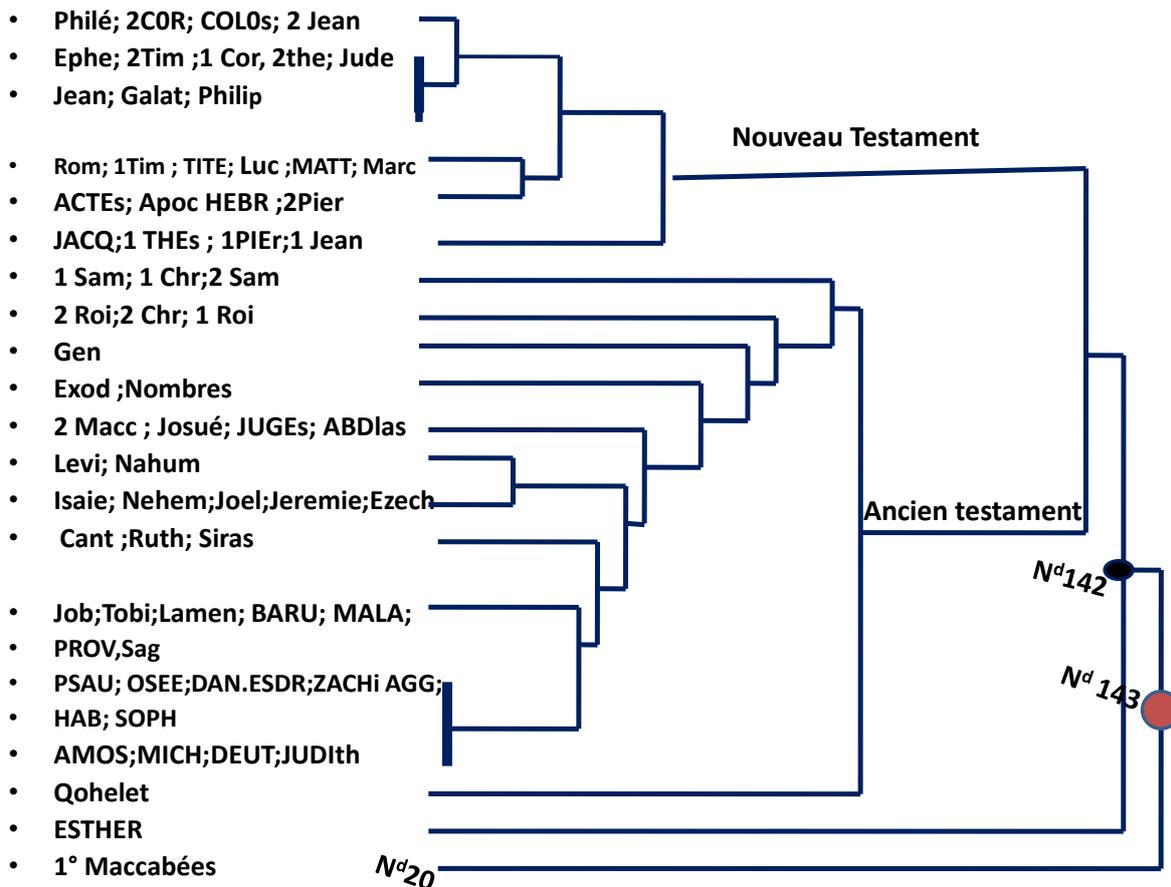
Nœud Général (N^d143) (fig. 8,9)

La première dichotomie située au haut de l'arbre, divise l'ensemble des livres en deux classes inégales : $N^d 143 = N^d 142(\text{Benjamin}) \cup N^d 20(\text{aîné : 1}^{\text{er}} \text{ Maccabées})$.

Nœud principal



(fig 8).



(fig.9) J.P.Benzécri & Kharchaf Idris, Colloque à Nice, « Méthode quantitative et informatique dans l'étude des textes », patronage du CNRS & ALLC, 5-8juin 1985, Nice, France.(Publié dans les Cahiers de l'Analyse des Données, Paris VI,France)

On a construit un tableau de correspondance, croisant avec l'ensemble des livres de la Bible les 520 mots ou locutions faisant l'objet du plus grand nombre de référence.

A l'intersection de la ligne m et de la colonne l , on a le nombre des références au livre l trouvées sous le mot m . Ce tableau a été soumis à l'analyse factorielle et à la classification automatique.

Nous nous bornons ici à présenter brièvement la classification obtenue sur l'ensemble des livres.

D'abord, une particularité : le 1^{er} livre des Maccabées et celui d'Esther se séparent de tout le reste de la Bible. C'est que dans la Table Pastorale, on a multiplié les références à certains personnages qui n'apparaissent que dans l'un ou l'autre de ces livres.

Ensuite, le Nouveau et l'Ancien Testament sont parfaitement séparés : on ne s'en étonnera pas non plus ; car chacun des mots de l'ensemble E= {Apostolat, Evangile, Foi, Jésus, Christ....} renvoie à la quasi-totalité des livres du N.T.

C'est sur le détail des subdivisions des deux parties de la Bible que nous appellerons l'attention des exégètes. Pourquoi, tandis que les Evangiles synoptiques {Luc, Matthieu, Marc} vont avec trois épîtres de S^t Paul {Romains, 1^{ère} à Timothée, Tite}, l'Evangile selon st Jean va-t-il avec deux autres épîtres {Gelâtes et Philippins} ?

Et dans l'A.T., comment expliquer la place du deutéronome, (dont le sigle DEUT est encadré comme ceux des autres livres du Pentateuque et des Evangiles) à proximité de nombreux livres prophétiques {dont les sigles sont surlignés) ainsi que Judith, de Daniel, d'Esdras et du Psautier ?

Commentaire du Professeur Jean Paul Benzecri (*)

L'Analyse des données peut expliquer comment ont été agrégées les classes de la CAH : il suffit en bref de croiser une partition des livres avec une partition des mots pour savoir (par des calculs précis) entre quelles classes de livres et quelles classes de mots

les références de la Table Pastorale ont établi les liens les plus étroites qui sont aussi la cause de la cohésion de ces classes.

Mais au fond, les références elles-mêmes, résultent d'un choix plus ou moins conscient de ceux qui ont conçu la Table Pastorale. Certes au niveau de chaque livre et de chaque mot, les exégètes ont, pas à pas, relevé ce qui leur semblait significatif. Mais le résultat, considéré dans son ensemble a une structure que révèle l'analyse des données. En résulte une vision du contenu de la Bible. Par-delà les erreurs de pondération (telles que celles qui ont donné au livre d'Esther un caractère unique qu'il n'a sans doute pas dans l'esprit des auteurs de la Table) n'est-ce pas la conception du message biblique sous-jacente à toute l'entreprise que nous découvrons ? Les exégètes peuvent maintenant regarder leur propre œuvre et faire un examen de conscience.

(*)(Professeur. J.P.Benzecri :D.2019)

[Ancien élève de l'École normale supérieure (1950) et fondateur de l'école française d'analyse des données dans les années 1960-1990, Jean-Paul Benzécri a développé des outils statistiques, notamment l'analyse factorielle des correspondances qui permet de traiter de grandes masses de données afin de visualiser et hiérarchiser l'information.

Il était professeur à la Faculté des sciences de Paris puis à l'Université Pierre-et-Marie-Curie, auxquelles l'Institut de statistique de l'université de Paris est successivement rattaché.

Avant de s'orienter vers les statistiques, il avait soutenu sous la direction d'Henri Cartan une thèse remarquable sur les variétés affines, où il démontrait que la caractéristique d'Euler-Poincaré d'une surface affine plate compacte est nulle. Wikipédia2019].

Etape3(E3) : Stylométrie

L'étude qui a été faite est basée sur la table pastorale*, qui constitue une présentation méthodique (en langue française) des mots, des idées, des sujets et des noms principaux du livre saint (ces derniers ont été pris selon leur ordre alphabétique, commençant par Aaron et terminant par Zakaria).

La table pastorale sert à retrouver dans la Bible tel ou tel passage à partir des mots principaux qu'il contient.

Donc les textes contenus dans la table pastorale sont de deux sortes :

- Les citations verbales, où se retrouve le mot-clef.
- Les citations idéologiques qui en contiennent l'idée.

Choix des mots

Soucieux d'élaborer les méthodes les plus efficaces pour comprendre la Bible, les méthodologistes (ou théologiens) sont conduits à apprendre le vocabulaire fondamental de la Bible.

C'est pourquoi il est intéressant de trier judicieusement le lexique élémentaire, c'est-à-dire de pouvoir connaître la structure essentielle dont le lecteur (débutant) aura le plus besoin pour effectuer des tâches simples :

a/ Lire facilement la Bible.

b/ Comparer la structure de la Bible à celle du Coran.

c/ Est-il possible d'étudier la formation d'un texte sur le plan de l'unité ou de l'hétérogénéité des sources à partir de la distribution des

éléments (mots, noms propres, verbes choisis) qui composent son discours? (Kharchaf Idris, « Stylométrie et théorie des sources pour les versets du livre de la Genèse », Cahiers de l'Analyse des Données, Vol XIII-1988-n°1-pp93-94, France).

Pour exploiter au mieux toute l'information contenue dans la Bible (en prenant en considération les tâches précédentes), on a introduit les mots sous leur forme principale.

I/ Prophètes : Aaron, Zakaria, Joseph, Jonas, Isaïe, Jacques, Salomon, etc.

II / Verbe : Allumer, appeler, appliquer, intimider, etc.

III/ Noms : Bonté, Puissance, Argile, etc.

IV/ Phrase : Dieu brise la force de l'ennemi, Dieu dirige les peuples, ...

V/ Noms de villes ou régions : Jérusalem, Bethleem,

Ainsi en examinant les classes des mots dont, la fréquence est au voisinage de 50 ou plus (comme le cas de création, Israël,...), on constate que nous pouvons réaliser une étude de comparaison avec le Coran dans différentes parties du texte.

Suggestions

Indépendamment des bases traditionnelles (travail de G. Weil, The genesis Books), concernant la nature des versets, nous allons proposer certaines modifications concernant les versets de quelques chapitres (selon la distance au sens de χ^2 et la projection sur un espace des facteurs F_α .(Tab.33A)

Donc il s'agit d'un tableau de la distribution des trois parties du discours hébreu de la Bible, sur un ensemble de 1533 versets (50 chapitres de la Genèse).

Les variables sont ainsi définies (Tab.17) :

Variables	Profil (Fréquence absolue)
Noms	10018
Verbes	4987
Particules	5613
BKLM (autres)	2171

(Tab.17)

Notation :

J : représente la source yahviste

E: représente les parties du récit où DIEU est désigné sous le nom d'Elohim.

Ces deux sources sont particulièrement apparentes dans les livres de la Genèse et de l'Exode.

Résultat : Analyse & représentation des graphiques

Le tableau global a pour dimension 1533*4, l'hypothèse posée pour l'étude de ce texte était de savoir dans quelle mesure il est possible d'étudier la formation d'un texte sur le plan de l'unité ou des sources à partir de la distribution des éléments qui composent son discours.

La question qui restait posée était de savoir quelle serait la nature et la dispersion des chapitres.

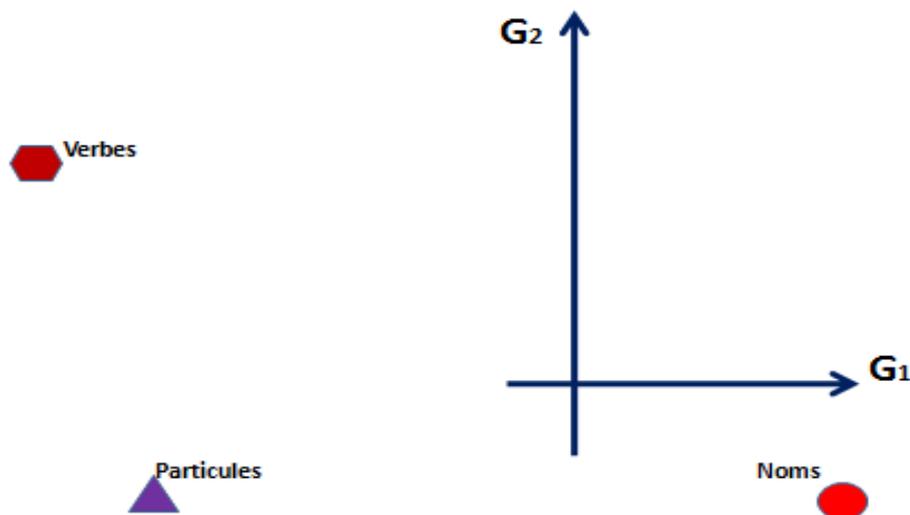
Résultats :

Les premières valeurs propres et les travaux d'inertie correspondants sont les suivantes (Tab.18) :

Les valeurs propres	$\lambda_1=0.118$	$\lambda_2=0.052$
Taux d'inertie	$\tau_1=69.453\%$	$\tau_2=30.547\%$

(Tab.18)

Sur le premier facteur se détache nettement du deuxième facteur (fig.10)



Projection sur l'espace $G_1 \times G_2$ (Fig.10)

Par rapport aux variables, nous constatons une séparation nette entre les noms d'un côté (G_1), et les {particules \cup Verbes} avec BKLM (en éléments supplémentaires).

Sur le deuxième facteur (G_2), l'opposition se passe entre les verbes et particules : $G(\text{verbes}) = +366$, $g(\text{particules}) = -264$;

$$\text{Ctr}(\text{verbes}, G_2) = 62.3\% ; \text{Ctr}(\text{particules}) = 36.7\%$$

Commentaire

Le tableau de la distribution des trois parties du discours hébreu de la Bible (cité ci-dessus), montre le suremploi généralisé des noms, et une tendance particulière et moins de verbes ou plus de verbes et moins de particules.

On constate donc, trois situations possibles:

1/ Les chapitres du livre marquent un suremploi des particules, et l'expérience montre que le style du livre est du type prose narrative (Exode pour le cantique de la mer (chap. 15), Juges (Chap. 5), Genèse (chap.4,5,11) et les nombres (chap. 7,10,34,35,36).

2/ Les chapitres du livre marquent un suremploi des verbes et le maximum des excroissances, l'expérience là aussi montre que le style du livre est poétique, lyrique, épique ou didactique ou encore il est rédigé en prose cadencée, qui est une des formes de la poésie hébraïque dans la Bible (Jes :36,39), (1 Rois :17-22), (2 Rois :2-9,10) sont bien marqués par un suremploi des verbes.

3/La situation est partagée et les chapitres du livre marquent tantôt un suremploi des particules, tantôt un suremploi des verbes.

L'expérience montre que le style du livre est du type prose narrative : comme le chapitre 15 de l'exode pour le cantique de la mer, la Genèse (Ch4, 5,11) et les nombres (7,10,34,35,36).

Pour le tableau :44 : cet ensemble est caractérisé par un suremploi de verbes 1Rois (17-22), 2 Rois (Ch2-9,10)

Projection des chapitres

Chapitre	Verset	Etiquetage initial	Etiquetage retouche
1	29	P	J
1	31	P	J
2	7	J	P
2	12	J	E
2	14	J	E
3	38	J	E
4	22	J	E
5	3	P	J
5	5	P	J
5	6	P	J
5	10	P	E
5	13	P	E
5	15	P	E
5	24	P	J
5	26	P	E
6	8	J	P
6	12	P	J
7	12	J	E
7	19	P	J
9	12	P	J
9	17	P	J
9	20	P	J
10	2	P	E
10	3	P	E
10	7	P	E
10	12	J	E
10	22	P	E
10	23	P	E
10	25	J	P

12	20	P	J
13	13	J	P
14	11	?	J
14	13	?	P
14	16	?	J
14	23	?	J
15	5	E ?	J
15	16	E ?	J
16	3	J	P
17	2	P	J
17	10	P	J
18	14	J	P
20	9	E	J
20	10	E	J
20	13	E	J
21	6	E	J
21	26	E	J
21	29	E	J
21	30	E	J
21	34	E	P
22	5	E	J
22	12	3	J
22	16	E	J
22	21	J	E
24	62	J	P
25	7	P	E
25	14	P	E
25	15	P	E
25	23	J	P
27	22	J	E
28	10	J	P
28	17 , 22	E	J

29	25	E	P
29	27	E	J
30	10	J	P
30	12	J	P
31	9	JUE	J
31	12	JUE	J
31	5	JUE	J
31	13	JUE	J
31	17	JUE	J
31	20, 21	JUE	J
31	32	JUE	J
31	42, 48, 50, 52	JUE	J
31	33	JUE	P
32	15	J	E
34	18	J	P
35	2	E	J
35	23, 25	P	P
36	6, 15, 19, 20, 41, 42	P	E
37	22	E	J
38	6, 7	J	P
39	1	J	E
40	8, 14, 16	E	J
41	22, 11, 37	E	P
41	24, 41, 43	E	J
42	3	E	P
42	23, 33, 36, 37	E	J
42	13	J	E
45	16	JUE	P
45	11, 15, 18	JUE	J
46	9, 11	P	E
48	8, 11, 21	JUE	J

49	14	J	E
49	11, 16, 24, 27	J	P
49	31	P	J
50	9	J U E	J
50	19, 21	E	J
50	22	E	P

Le tableau qui représente la correction de la nature des chapitres de la Bible (Tab.19)

Bibliographie

1. Bible et informatique : le texte (Louvain-la-Neuve, 2-4 septembre 1985)
2. [Compte-rendu] Association internationale « Bible et Informatique : Le texte » Louvain-la-Neuve, en collaboration avec la Faculté de théologie, B-5198 Denée, abbaye de Maredsus.
3. Actes du Quatrième Colloque international Bible et informatique : Matériel et matière : l'impact de l'informatique sur les études bibliques ... Bible studies : Amsterdam, 15-18 August 1994 Broché – 19 septembre 1995
4. Benzécri, J.-P. L'analyse des données, tome 1, La taxinomie, tome 2, L'analyse des correspondances. Dunod, Paris, 1973.
5. Crucianu, M., J.-P. Asselin de Beauville, et R. Boné. Méthodes d'analyse factorielle des données : méthodes linéaires et non linéaires. Hermès, Paris, 2004.
6. Saporta, G. Probabilités, Analyse des Données et Statistique. Technip, Paris, 2011.
7. Kharchaf Idris, Probabilité Mathématique, et initiation à l'Analyse des données, collection Al Hissab, KWD, 1992, rabat, maroc

8. Blanchard Philippe, Patou Charles, « Les usages de l'analyse factorielle dans les revues de sciences sociales »,
9. Méthodes et outils des sciences sociales. Innovation et Renouveau, L'harmattan, collection **Cahiers Politiques**, Paris, 2003, p. 85-110.
10. CREDEP - Université Paris 9 - Dauphine

+++++