

A. CHABIR

**Adjonction de textes arabes à un corpus de base  
par l'analyse discriminante et la classification  
ascendante hiérarchique**

*Les cahiers de l'analyse des données*, tome 22, n° 3 (1997),  
p. 319-330

[http://www.numdam.org/item?id=CAD\\_1997\\_\\_22\\_3\\_319\\_0](http://www.numdam.org/item?id=CAD_1997__22_3_319_0)

© Les cahiers de l'analyse des données, Dunod, 1997, tous droits réservés.  
L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# ADJONCTION DE TEXTES ARABES À UN CORPUS DE BASE PAR L'ANALYSE DISCRIMINANTE ET LA CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

[DISCR. ARABE]

A. CHABIR\*

## 1 Corpus de base et textes nouvellement saisis

L'article [MOTS ARABES] (in *CAD*, Vol XIX, n°1) rend compte d'une expérience stylométrique portant sur 13 fragments de textes longs, chacun, d'environ 3000 caractères (3k); et tous compris dans le domaine de la philosophie, même s'ils diffèrent quant au niveau de l'exposé.

Dans [TEXTES ARABES] (in *CAD*, Vol XX, n°4; cité, dans la suite: [TxAr]), est traité, suivant la même méthode, un corpus étendu, où, aux 13 premiers fragments sont ajoutés 78 autres, longs également de quelque 3k, mais d'une plus grande diversité de genre: la cueillette, sans prétendre aucunement embrasser l'infinie diversité de ce qu'on peut lire en arabe, sort de la philosophie pour glaner dans la prose et la poésie classique; et ne fait pas fi de ce qu'ont produit les modernes; en littérature proprement dite, et aussi dans la presse quotidienne.

À l'occasion de notre thèse, consacrée aux *Séances d' al Hamadānī*, nous avons considéré dans le détail, non seulement le texte de chaque maqāmā, مقامة, mais aussi le texte intégral d'un autre chef-d'œuvre: kalīlah wa dimnah, كليله ودمنه, où ibn-l-muqaffa<sup>4</sup>, ابن المقفع, suivant la tradition de l'Inde et de l'Iran, met en arabe des fables dont héritera LA FONTAINE.

Ces deux classiques sont déjà représentés dans le corpus de [TxAr], le premier par quatre fragments, {\$Mqm, mqmB, mqm1, mqm2}; le second par un seul: \$KwD. Complètement saisis, ils constituent, respectivement, 51 et 37 fragments. Le volume des textes se trouve ainsi doublé.

Mais nous ne présenterons pas une analyse factorielle globale d'un tel corpus, où prédomineraient deux grandes œuvres, certes séparées dans le

---

(\*) Maître de langue à l'Institut National des Langues et Civilisations Orientales.

temps par deux siècles, mais relativement proches quant au style. Nous préférons conserver tel quel le cadre multidimensionnel créé, dans [TxAr], et y insérer, en éléments supplémentaires, les (51+37) nouveaux fragments.

De façon précise, on a repris, du §4.2 de [TxAr], l'analyse de la correspondance entre le lexique étendu:  $\Delta 5$ , de 232 formes (ou locutions); et 89 fragments de base, la poésie moderne, {occ, dpl}, étant écartée; et la prose moderne, {Grb, Sgn, Prs}, repondérée, comme on l'explique dans [TxAr], §2.4.

Les (51+37) fragments, ont été dépouillés, comme ceux du corpus de base (cf. [TxAr], §§2.1, 2.2, 2.3; et [MOTS ARABES]); d'où un tableau (232)  $\times$  (51+37), adjoint en supplément à l'analyse du tableau de base: (232)  $\times$  (89). Ainsi, l'on a pour chacun des (51+37) fragments, des coordonnées sur les axes factoriels issus de l'analyse de base.

Dans la suite, on se placera dans l'espace engendré par les 20 ou les 50 premiers axes. En termes géométriques, dans un tel espace, chaque fragment – fragment de base ou fragment supplémentaire – est représenté par un point. Entre deux points, est définie une distance, analogue multidimensionnelle de la distance de l'espace euclidien usuel.

Désignons par  $J_1$ ,  $J_2$  des ensembles finis de points. Par le programme: *discri*, d'analyse discriminante, on détermine, pour chaque point  $j_2$  de  $J_2$ , le point  $j_1$  qui, de tous les points de  $J_1$ , est le plus proche de  $j_2$ : on dit alors que  $j_2$  est affecté à  $j_1$ ; expression qui s'explique parce que, si l'on connaît l'auteur, ou seulement le genre, des individus de  $J_1$  mais non de ceux de  $J_2$ , l'assimilation à  $j_1$  offre une présomption quant à ce qu'est  $j_2$ .

Selon la méthode présentée dans [DISCR. CAH LING.], avec son application à des textes grecs, l'étude des (51+37) fragments sera faite ici par le programme '*discri*' ainsi que par le programme de classification '*CLH*', déjà utilisé, dans [TxAr], pour le corpus de base.

De façon précise, on désignera désormais par:

$J$  : l'ensemble des 89 fragments du corpus de base, décrit, avec les sigles adoptés, dans [TxAr] ;

$Q_b$  : l'ensemble des 51 maqâmâ, avec pour sigles: {mq01, mq02, ..., mq50, mq51} ;

$KDb$  : l'ensemble des 37 fragments de: kalîla wa dimnâ, dont les sigles : {kdAa, ..., kdAh, kdBa, kdCa, ..., kdCe, ..., kdOb, kdPa} suivent une division hiérarchique des parties du texte en fragments.

Éventuellement, on considérera la réunion de deux ou de trois des ensembles  $J$ ,  $Q_b$ ,  $KDb$ .

## 2 Applications de l'analyse discriminante

### 2.1 Affectation des maqâmât au corpus de base

Nous considérerons deux affectations de Qb à J : selon que l'espace ambiant est engendré par 20 ou 50 axes.

#### 2.1.1 Affectation dans l'espace de dimension 20

Vu la diversité du corpus de base, où les maqâmât ne figurent que par 4 fragments, l'expérience acquise dans d'autres études suggère qu'il vaut mieux ne pas prendre en compte toutes les dimensions mais se borner à la vue globale offerte par un nombre modéré de facteurs: ainsi 20 axes rendent compte de 57% de l'inertie du nuage J des 89 fragments de base.

Sans: Dkal:C&MΔ5\$corA& ; nombre de facteurs utilisés = 20

affectation des qb aux j (corpus de base)

(mq01->\$Mqm) (mq02->7Sgn) (mq03->mqm1) (mq04->mqm1) (mq05->7Sgn) (mq06->bSgn)  
 (mq07->flshu) (mq08->3Sgn) (mq09->8Sgn) (mq10->7Sgn) (mq11->dSgn) (mq12->mqmB)  
 (mq13->mqm1) (mq14->mqm1) (mq15->5Grb) (mq16->flshu) (mq17->4Grb) (mq18->mqm1)  
 (mq19->eSgn) (mq20->7Sgn) (mq21->\$twh) (mq22->mqm2) (mq23->\$twh) (mq24->4Grb)  
 (mq25->6Sgn) (mq26->4Grb) (mq27->6Sgn) (mq28->\$tah) (mq29->πsc2) (mq30->aPrs)  
 (mq31->TfGh) (mq32->\$twh) (mq33->mqm1) (mq34->mqm1) (mq35->6Sgn) (mq36->flshu)  
 (mq37->\$twh) (mq38->\$twh) (mq39->3Sgn) (mq40->\$twh) (mq41->2Adb) (mq42->dSgn)  
 (mq43->mqm2) (mq44->\$twh) (mq45->\$twh) (mq46->1Sgn) (mq47->\$twh) (mq48->flshu)  
 (mq49->7Sgn) (mq50->dSgn) (mq51->6Sgn)

On a rappelé que J comprend 4 fragments des maqâmât: à \$Mqm et à mqmB, fragments qui coïncident respectivement avec mq01 et mq12, il n'y a qu'une seule affectation, celle attendue. Le cas de {mqm1, mqm2}, moitiés initiale et finale de mq22, est différent: il y a 7 affectations à mqm1; et 2 à mqm2: d'une part, la maqâmâ complète: mq22; et d'autre part: mq43.

La distance de mq43 à mqm2 est toutefois très grande, dépassant 10 fois les autres distances d'affectation trouvées ici: car mq43 ne ressemble, de près ni de loin, à rien d'autre. L'éditeur, Muḥammad ʿabduh, a renoncé à en publier le texte intégral. Deux loustics, s'affrontant dans un concours d'insultes, dont l'enjeu est un dinar: d'où le titre: al-maqāmah al-dīnāriyyah. L'essentiel des répliques consiste en une particule du vocatif: yā, suivie d'un nom et de son complément; le tout ne pouvant avoir qu'un profil excentrique sur l'ensemble des mots outil dénombrés.

Un autre morceau singulier est mq30: al-maqāmah al-ruṣāfiyyah, elle aussi censurée par l'éditeur. Aux abords du lieu de Bagdad qui donne son nom à mq30, des désœuvrés, ayant fui, dans une mosquée, la chaleur torride des rues, trompent leur ennui en énumérant les procédés des voleurs. Ce, dans une suite de formules, à peine moins concises que celles de mq43. L'affectation est faite à aPrs (éditoriaux de presse moderne): la distance de mq30 à aPrs, est moins du tiers de celle séparant mq43 de mqm2; mais toutes les autres distances d'affectation sont moins de la moitié de (mq30, aPrs).

On ne s'étonnera pas que 9 maqâmât soient affectées à \$twh: "marque de la sollicitude d'un calife envers la plèbe de Bagdad".

Mais nombre d'affectations vont à des modernes: 4 à flshu: le *Prince des poètes*, Aḥmad Ṣaūqī, أحمد شوقي, (1868-1932); 4 à Grb, fragments du

livre *al-ġirbāl*, الغربال, - le tamis - : théorie des belles lettres proposée par Mīkātīl Nu'aymah, ميخائيل نعيمة, auteur libanais mort quasi centenaire en 1988; et 18 à Sgn, *siġn al-ʿumr*, سجن العمر : la prison des jours, (littéralement: de l'âge), de: Tawfīq al-Hākīm, توفيق الحكيم, romancier égyptien du XX-ème siècle.

### 2.1.2 Affectation dans l'espace de dimension 50

Sans:DKal:C&MΔ5ScorA& ; nombre de facteurs utilisés = 50

affectation des qb aux j (corpus de base)

(mq01->SMqm) (mq02->7Sgn) (mq03->7Sgn) (mq04->7Sgn) (mq05->7Sgn) (mq06->gGrb)  
 (mq07->3Grb) (mq08->7Sgn) (mq09->aSgn) (mq10->7Sgn) (mq11->6Grb) (mq12->mqmB)  
 (mq13->6Grb) (mq14->7Sgn) (mq15->7Grb) (mq16->3Grb) (mq17->gGrb) (mq18->gGrb)  
 (mq19->dSgn) (mq20->bSgn) (mq21->6Grb) (mq22->mqm2) (mq23->iGrb) (mq24->6Grb)  
 (mq25->bSgn) (mq26->iGrb) (mq27->6Sgn) (mq28->gGrb) (mq29->SMqm) (mq30->mqm1)  
 (mq31->2Adb) (mq32->7Sgn) (mq33->7Sgn) (mq34->mqm1) (mq35->7Sgn) (mq36->kGrb)  
 (mq37->kGrb) (mq38->2Adb) (mq39->6Grb) (mq40->iGrb) (mq41->2Adb) (mq42->7Sgn)  
 (mq43->mqm1) (mq44->2Grb) (mq45->Stwh) (mq46->Iaπ) (mq47->dSgn) (mq48->dSgn)  
 (mq49->7Sgn) (mq50->2Adb) (mq51->7Sgn)

Les axes 1 à 50 rendent compte de plus de 88% de l'inertie du nuage J.

À ce niveau de précision, il a d'abord 7 affectations aux fragments de maqâmā de J; celles, attendues et déjà vues au §2.1.1, de mq01, mq12, mq22; mais aussi, les censurées: mq43 et mq30; avec, ici encore, des distances maxima.

Dans le domaine classique, on note 4 affectations à: *al-ʿadāb al-kabīr*; œuvre d'ibn-l-Muqaffaʿ, comme *kalīlah wa dimnah*.

Mais prédominent les affectations à deux modernes déjà cités: 18 à Grb et 20 à Sgn.

Destinés à plaire et à convaincre, traitant divers thèmes se référant aux valeurs modernes, respectant la pureté de la langue, tout.. en se plaçant comme néo-classiques ou novateurs, Grb et Sgn tirent leur force expressive du cadre narratif classique habilement amélioré par *Ḥamaḡḡānī*, cadreoù prédominent:

Les particules exclamatives ou interrogatives qui appellent l'attention du lecteur ou de l'auditeur.

Les innombrables apostrophes formées avec: Ya.

Les conjonctions de subordination servant à introduire des précisions, des explications, d'où la fréquence des phrases brèves juxtaposées.

Les pronoms personnels et les pronoms relatifs, dus, en grande partie, à l'alternance entre récit et discours.

Les particules du cas direct pour mettre en relief un pôle ou un actant.

### 2.2 Affectation des maqâmāt au corpus de base, complété par les 37 fragments de kalīlā wa dimnā

Bien qu'aucune affectation n'ait été trouvée à §KwD, unique fragment représentant *و دمنة كليله* dans le corpus de base J, l'expérience devait être

tentée d'affecter Qb à la réunion:  $JKb = J \cup KDb$ . Il apparaît alors que, par la vivacité et la fluidité de son style, kalilä wa dimnä approche les maqâmât de si près, que les modernes ne reçoivent presque aucune affectation.

Sans:DKal:C&MΔ5\$corA& ; nombre de facteurs utilisés = 20

affectation des qb aux jkb (corpus de base et maqâmât)

(mq01->SMqm) (mq02->kdPa) (mq03->kdPa) (mq04->kdCe) (mq05->kdPa) (mq06->kdCe)  
 (mq07->kdCa) (mq08->kdPa) (mq09->8Sgn) (mq10->kdFa) (mq11->kdCb) (mq12->mqmB)  
 (mq13->kdFa) (mq14->kdPa) (mq15->kdAe) (mq16->kdPa) (mq17->kdCe) (mq18->kdPa)  
 (mq19->kdEb) (mq20->7Sgn) (mq21->kdFa) (mq22->mqm2) (mq23->kdBa) (mq24->kdCd)  
 (mq25->kdCe) (mq26->kdCa) (mq27->6Sgn) (mq28->\$tah) (mq29->πsc2) (mq30->aPrs)  
 (mq31->kdOa) (mq32->kdPa) (mq33->kdPa) (mq34->mqm1) (mq35->kdPa) (mq36->kdPa)  
 (mq37->kdAf) (mq38->kdCc) (mq39->kdAe) (mq40->kdFa) (mq41->kdPa) (mq42->kdJc)  
 (mq43->mqm2) (mq44->kdCe) (mq45->kdEa) (mq46->1Sgn) (mq47->kdPa) (mq48->kdPa)  
 (mq49->7Sgn) (mq50->kdCa) (mq51->kdEb)

Dans l'espace rapporté aux 20 premiers axes, on a 38 affectations de maqâmât à كليلة ودمنة : kd; 5 aux quatre fragments de maqâmä qui sont dans J; 5, seulement, à الغريال (Grb: le tamis).

bilan des affectations des maqâmât

|              | KwaD | mqm | Grb | Sgn | autres |
|--------------|------|-----|-----|-----|--------|
| avec 20 axes | 38   | 5   | 5   | 0   | 3      |
| avec 50 axes | 41   | 7   | 1   | 2   | 0      |

Sans:DKal:C&MΔ5\$corA& ; nombre de facteurs utilisés = 50

affectation des qb aux jkb (corpus de base et maqâmât)

(mq01->SMqm) (mq02->kdCb) (mq03->kdGa) (mq04->kdCe) (mq05->kdPa) (mq06->kdCa)  
 (mq07->kdKd) (mq08->kdOb) (mq09->aSgn) (mq10->kdFa) (mq11->kdCa) (mq12->mqmB)  
 (mq13->kdCe) (mq14->kdCe) (mq15->7Grb) (mq16->kdGa) (mq17->kdCe) (mq18->kdAe)  
 (mq19->kdAa) (mq20->bSgn) (mq21->kdNa) (mq22->mqm2) (mq23->kdBa) (mq24->kdAh)  
 (mq25->kdCe) (mq26->kdBa) (mq27->kdPa) (mq28->kdKc) (mq29->SMqm) (mq30->mqm1)  
 (mq31->kdCa) (mq32->kdCe) (mq33->kdCa) (mq34->mqm1) (mq35->kdCa) (mq36->kdCa)  
 (mq37->kdAf) (mq38->kdCa) (mq39->kdAe) (mq40->kdCe) (mq41->kdKa) (mq42->kdKd)  
 (mq43->mqm1) (mq44->kdCe) (mq45->kdBa) (mq46->kdEa) (mq47->kdAh) (mq48->kdAg)  
 (mq49->kdJc) (mq50->kdOb) (mq51->kdCa)

Dans l'espace rapporté aux 50 premiers axes, les affectations à kalilä wa dimnä sont encore plus nombreuses: 41; il y en a 7 aux fragments de maqâmä; et 3 autres à Grb ou Sgn (la prison des jours).

### 2.3 Affectation des fragments de kalilä wa dimnä au corpus de base

sans:DKal:C&MΔ5\$corA& ; nombre de facteurs utilisés = 20

affectation des kdb aux j (corpus de base)

(kdAa->4Rsl) (kdAb->\$twh) (kdAc->2Adb) (kdAd->\$twh) (kdAe->\$twh) (kdAf->\$twh)  
 (kdAg->2Adb) (kdAh->\$twh) (kdBa->\$twh) (kdCa->dSgn) (kdCa->\$twh) (kdCb->\$twh)  
 (kdCc->\$twh) (kdCd->\$twh) (kdCe->dSgn) (kdEa->\$twh) (kdEb->6Sgn) (kdFa->\$twh)  
 (kdGa->2Adb) (kdGb->2Adb) (kdHa->2Rsl) (kdHb->2Rsl) (kdIa->4Rsl) (kdIb->2Rsl)  
 (kdJa->dSgn) (kdJb->dSgn) (kdJc->dSgn) (kdKa->2Adb) (kdKb->2Adb) (kdKc->2Adb)  
 (kdKd->dSgn) (kdLa->\$twh) (kdMa->4Rsl) (kdNa->\$twh) (kdOa->4Rsl) (kdOb->7Sgn)  
 (kdPa->\$twh)

Dans l'espace engendré par 20 axes, quinze affectations vont à \$twh: "calife et plèbe de Bagdad"; 7 à Rsl et 7 à Adb, œuvres qui sont du même auteur que k. wa d. À quoi s'ajoutent 8 affectations à Sgn.

Avec 50 facteurs, on a 19 affectations à Adb, une à Rsl; 3 à Grb; 14 à Sgn.

La manière de l'auteur est bien reconnue d'après le profil d'emploi des mots outil; mais, parmi les modernes, Sgn est proche de ce même profil.

### 3 Applications de la classification ascendante hiérarchique

#### 3.1 Classification des (51+37) fragments de deux œuvres

| c   | Partition en 14 classes : Sigles des fragments de la classe numéro c  |
|-----|---|
| 43  | mq43  |
| 30  | mq30  |
| 12  | mq12  |
| 22  | mq22  |
| 9   | mq09  |
| 162 | mq06 mq04 mq49 mq10 mq34 mq21 mq13 mq18 mq40 mq24 mq17 mq27 mq33 mq03 |
|     | mq35 mq25 mq32 mq51 mq20 mq08 mq39 mq07 mq05                          |
| 161 | mq36 mq48 mq16 mq14 mq37 mq47 mq41 mq26 mq23 mq15 mq11                |
| 142 | mq29 mq01   |
| 158 | kdKa kdOb kdLa kdBa   |
| 157 | mq46 mq02 mq45 kdPa kdAe kdFa kdEa kdCb kdCa kdJa kdJc                |
| 156 | mq42 mq31 kdOa mq50 mq19 kdKd kdCa                                    |
| 153 | kdEb kdJb kdCe kdAf kdKc kdNa mq44                                    |
| 159 | kdAh kdIb mq38 kdCc kdAa kdAc kdGa kdMa kdIa kdAb kdAg kdKb kdAd kdCd |
| 160 | kdGb kdHa kdHb mq28   |

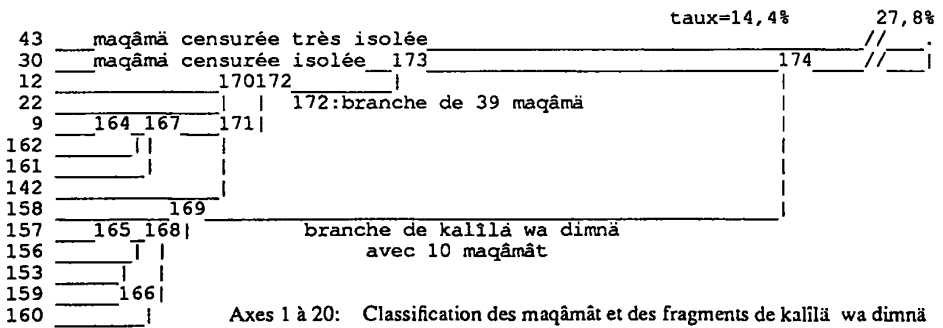
Axes 1 à 20: CAH: maqâmât et fragments de kalilâ wa dimnâ

Les analyses discriminantes du §2.2 ont montré une affinité particulière entre les maqâmât, d'une part, et kalilâ wa dimnâ, de l'autre. En soumettant à la classification l'ensemble des (51+37) fragments en lesquels ont été découpées ces deux œuvres, nous cherchons si cette affinité résulte d'un recouvrement réciproque ou d'une simple proximité.

En CAH, comme en analyse discriminante, on procède dans l'espace engendré soit par 20, soit par 50 axes.

Dans les deux classifications obtenues, la maqâmâ censurée mq43 se sépare des autres fragments au sommet de la hiérarchie.

Le reste se partage en deux branches: l'une composée exclusivement de maqâmât; l'autre qui, avec la totalité des 37 fragments de k wa d, comprend

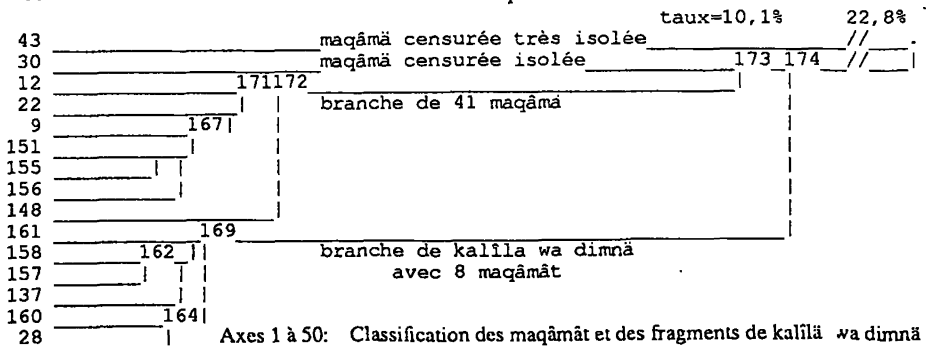


| c   | Partition en 15 classes : Sigles des fragments de la classe c                    |
|-----|--|
| 43  | mq43   |
| 30  | mq30   |
| 12  | mq12   |
| 22  | mq22   |
| 9   | mq09   |
| 151 | mq34 mq06 mq36 mq47 mq07 mq48 mq16 mq51 mq10 mq21 mq19                           |
| 155 | mq27 mq08 mq13 mq35 mq03 mq32 mq14 mq04 mq20 mq40 mq39 mq18 mq25 mq24            |
| 156 | mq37 mq17 mq44   |
| 156 | mq49 mq02 mq33 mq11 mq15 mq45 mq26 mq23  |
| 148 | mq29 mq01  |
| 161 | mq46 mq42 kdOa kdCa kdJa kdJc  |
| 158 | mq05 kdPa mq50 kdOb kdLa kdKc kdBa   |
| 157 | kdKd kdCa kdEb kdJb kdAa kdAc  |
| 137 | kdKa mq31 mq41   |
| 160 | kdGa kdAg kdEa kdCe kdAf kdAe kdAd kdFa kdNa kdCb kdMa kdAh kdCd kdKb            |
| 160 | mq38 kdCc kdIb kdIa kdAb kdGb kdHa kdHb  |
| 28  | mq28 Axes 1 à 50: Classification des maqâmât et des fragments de kalilâ wa dimnâ |

les quelques maqâmât absentes de la première branche. La maqâmâ censurée mq30 occupe, au sein de celle-ci, une position écartée; particulièrement dans la CAH fondée sur 50 facteurs.

Les maqâmâ {mq12, mq22}, qu'on a saisies d'abord dans le corpus de base parce qu'elles sont de toutes les anthologies, se distinguent. Se distingue également mq01 (apophtegmes sur les poètes arabes), texte tout aussi classique; et associé ici à mq29, où, en formules lapidaires, un éloquent expert décrit les perfections d'un cheval, devant l'illustre سيف الدولة الحمداني, saïf aldaulâ, seigneur d'Alep (915-965), prince également généreux dans la guerre et dans la paix.

On conclura que la séparation des œuvres est assez claire; les deux classifications s'accordant même entre elles quant aux détails.





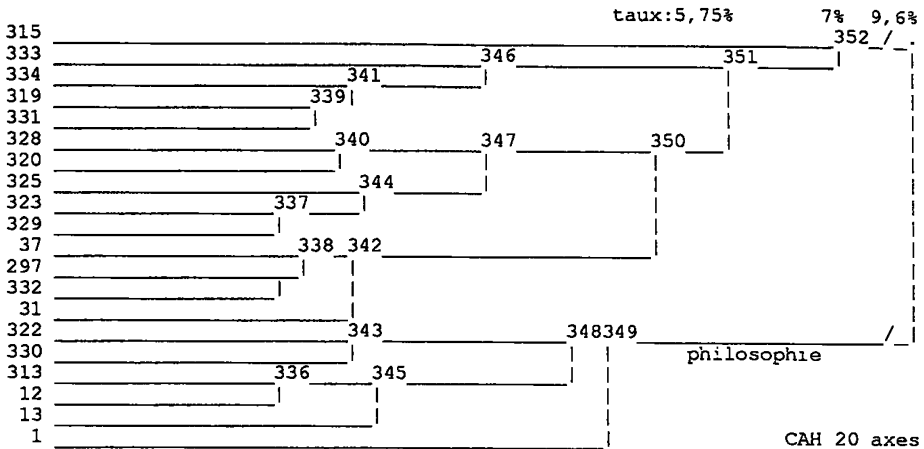
**3.2 Classification de (89+51+37) fragments, du corpus de base et de deux œuvres**

**3.2.1 Classification dans l'espace de dimension 20**

Les cohésion des deux œuvres dont le texte intégral est considéré dans le présent article, invite à les incorporer à une classification générale, avec les 89 fragments du corpus de base déjà pris dans leur ensemble au §4.2 de [TEXTES ARABES].

Dans l'espace engendré par les 20 premiers axes, la CAH sépare d'abord une branche 349 groupant les textes de philosophie. Le fragment apologétique chrétien, isf†, est à part. Le commentaire d'Aristote, par Ibn Roshd, exposé dont la rigueur logique ne laisse rien à désirer, est bien distingué et forme la classe 345. Restent dans la classe 343 des fragments dont le genre relève de la dialectique plutôt que de la stricte démonstration.

| c   | Partition en 20 classes : Sigles des fragments de la classe numéro c |                                  |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
|-----|--|----------------------------------|-----------------------------------|---------------------------------------|---------------------------|-------|--------------------|-------|------|------|---------|------|------|------|------|-----------------|------|------|--------------|------|------|------|------|------|------|------|------|------|
| 315 | 1Prs   | 2Prs                             | 6Prs                              | 4Prs                                  | 5Prs                      | 7Prs  | Presse (politique) |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 333 | mq43 (censurée)  | mq22                             | mqm2                              |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 334 | mqmB   | mq12                             | mqm1                              | ¶shu                                  | mq06                      | mq04  | mq49               | mq10  | mq34 | mq21 | mq13    | mq18 | mq40 | mq24 | mq17 | mq27            | mq33 | mq03 | mq35         | mq25 | mq32 | mq51 | mq20 | mq08 | mq39 | mq07 | mq05 | mq36 |
|     | mq48   | mq16                             | mq14                              | mq37                                  | mq47                      | mq41  | mq26               | mq23  | mq15 | mq11 | maqâmât |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 319 | SMqm   | mq01                             | πsc4                              | πsc5                                  | mq29                      | πsc2  | poésie             |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 331 | πsc3   | ¶Lac                             | πsc1                              | (y compris: (Mqm=mq01))               |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 328 | mq30 (censurée)  | aPrs                             | 8Prs                              | 9Prs                                  | 3Prs                      | bPrs  | Presse             |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 320 | &Id¶   | &Id\$                            | &Idπ (notices)                    | ¶Arb                                  | \$KWd                     | \$rbh | TfGh               |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 325 | aSgn   | 6Sgn                             | 7Sgn                              | bSgn                                  | 3Sgn                      | 1Sgn  | 4Sgn               | mq09  | 8Sgn | 5Sgn | §jhz    |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 323 | ¶jbr   | aGrb                             | Sgn: la prison des jours          |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 329 | cSgn   | 2Sgn                             | eSgn                              | dSgn                                  | kdJa                      | kdJc  | 9Sgn               | \$tah | mq28 | bGrb | 7Grb    | cGrb | dGrb | 8Grb | 9Grb | hGrb            | 5Grb | 6Grb | kdCa         | eGrb | kGrb | 4Grb | 3Grb | jGrb | iGrb | 2Grb | gGrb | 1Grb |
|     | fGrb   | Grb: le tamis (critique)         |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 37  | 1Adb   |                                  |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 297 | mq46   | mq02                             | mq45                              | kdPa                                  | kdAe                      | kdFa  | kdEa               | kdCb  | kdEb | kdJb | kdCe    | kdAf | kdKc | kdNa | mq44 | mq42            | mq31 | kdOa | mq50         | mq19 | kdKd | kdCa | kdKa | kdOb | kdLa | kdBa | kdAh | kdIb |
|     | mq38   | kdCc                             | kdAa                              | kdAc                                  | kdGa                      | kdMa  | kdIa               | kdAb  | kdAg | kdKb | kdAd    | kdCd | kdGb | kdHa | kdHb | kalîlâ wa dîmnâ |      |      |              |      |      |      |      |      |      |      |      |      |
| 332 | 3Rsl   | 1Rsl                             | \$Hdt                             | \$twh                                 | 2Rsl                      | 4Rsl  | 2Adb               | 4Adb  | 3Adb |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 31  | \$sfa  |                                  |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 322 | TfHy   | Rmq1                             | Rmq2                              | Rmq3                                  | Ibn Roshd: traité décisif |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 330 | ghz2   | ghz3                             | ghz1                              | Ghazali: effondrement des philosophes |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 313 | rZt1   | rZt2                             | Ibn Roshd: commentaire d'Aristote |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 12  | rZt3   | (livre Z de la métaphysique)     |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 13  | rZt4   |                                  |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      |              |      |      |      |      |      |      |      |      |      |
| 1   | isf†   | Épître sur l'Unité et la Trinité |                                   |                                       |                           |       |                    |       |      |      |         |      |      |      |      |                 |      |      | CAH: 20 axes |      |      |      |      |      |      |      |      |      |



Des deux fragments pris dans le "philosophe autodidacte" d'Ibn Tufayl, seul TfHy, qui relève de la théologie, est dans la branche 369; l'autre fragment, TfGh, consacré à Ghazâlî, est ailleurs: et ne s'en étonnera pas qui a présent à l'esprit la vivacité du texte, portrait d'un philosophe insaisissable, autant et plus qu'exposé des pensées de celui-ci.

Au sein du reste, qui constitue la branche 352, la classe 315, qui comprend six fragments de presse politique, est isolée.

La masse principale des textes littéraires est dans la branche 351; qu'on considérera en trois divisions: 346, 347, 342.

Dans 346, sont la plupart des maqâmât, avec les fragments poétiques du corpus de base (cf. TxAr, §1.1).

L'essentiel de 347 comprend Sgn - la prison des jours; et Grb - le tamis. En distinguant dans la classe 329 une subdivision (renfermée entre parenthèses dans le tableau du contenu des classes), on sépare à peu près ces deux œuvres modernes; au sein des quelles sont insérés quelques fragments classiques.

Dans 342 sont la plupart des kalîlâ wa dimnâ, avec ceux du même auteur (ابن المقفع) compris dans le corpus de base: Adb (الأدب الكبير), et Rsl (رسالة في الصحابة).

Quant au détail, on note que, comme dans la CAH du §4.2 de TxAr, la poésie f|jbr, de Gibran (جبران خليل جبران), est proche de Grb; et plus particulièrement du fragment aGrb. Qu'entre les classes 315 et 320 de la présente CAH, les fragments de Presse se partagent comme au §4.2 de TxAr (entre 136 et 164); les notices littéraires, &Id, étant, dans les deux CAH, associées aux mêmes fragments de presse (aPrs, bPrs...).

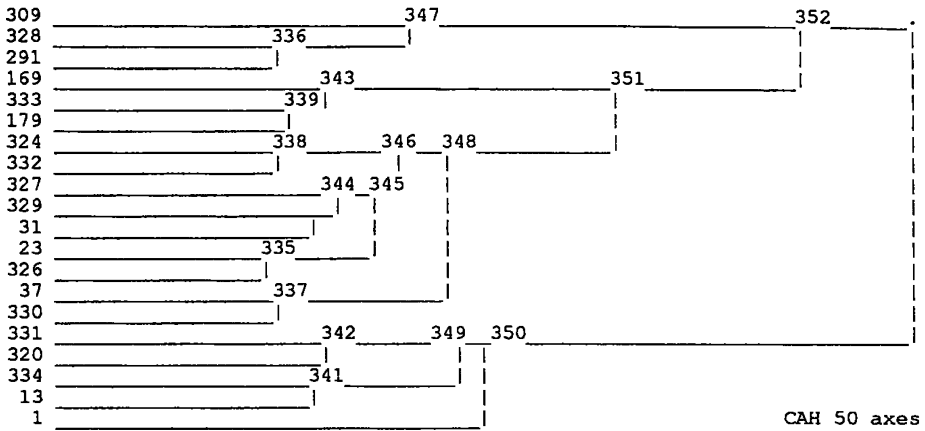
### 3.2.2 Classification de (89+51+37) fragments dans l'espace de dimension 50

| c   | Partition en 20 classes : Sigles des fragments de la classe numéro c  |
|-----|---|
| 309 | 6Prs 1Prs 2Prs 5Prs 7Prs Presse   |
| 328 | aPrs 9Prs 8Prs 4Prs 3Prs bPrs   |
| 291 | &Id¶ &Id\$ &Idπ notices   |
| 169 | mq43 (censurée)   |
| 333 | mq30 (censurée) mqm1 mq22 mqm2  |
| 179 | (mqmB mq12)   |
| 324 | (\$Mqm mq01)mq29 πsc4 πsc2 πsc5 Poésie  |
| 332 | πsc3 ¶Lac πsc1  |
| 327 | 6Sgn aSgn mq09 bSgn 7Sgn 5Sgn 3Sgn 8Sgn 1Sgn 2Sgn cSgn 4Sgn 5jhz Sgn: la prison des jours   |
| 329 | (¶shu mq34 mq06 mq36 mq47 mq07 mq48 mq16 mq51 mq'0 mq21 mq19 mq27 mq08<br>mq13 mq35 mq03 mq32 mq14 mq04 mq20 mq40 mq39 mq18 mq25 mq24 mq37 mq17<br>mq44 mq49 mq02 mq33 mq11 mq45 mq26 mq23 ¶Arb \$Hdt) maqâmât<br>\$KwD \$twh kalîlâ wa dimnâ<br>(eSgn mq46 mq42 kdOa kdJa kdJc dSgn mq05 kdPa mq50 kdOb kdLa kdKc kdBa<br>kdKd kdCa kdEb kdJb kdAa kdAc kdGa kdAg kdEa kdCe kdAf kdAe kdAd kdFa<br>kdNa kdCb kdMa kdAh kdCd kdKb mq38 kdCc kdIb kdIa kdAb kdGb kdHa kdHb<br>(\$rbh TfGh 9Sgn \$tah TfHy) ibn tufayl séparé du reste de la philosophie<br>31) \$sfa |
| 23  | ¶jbr  |
| 326 | aGrb mq15 7Grb bGrb cGrb dGrb 8Grb eGrb 4Grb 9Grb 5Grb 6Grb kdCa jGrb<br>hGrb fGrb gGrb iGrb 3Grb kGrb lGrb 2Grb mq28 Grb: le tamis (critique)  |
| 37  | lAdb  |
| 330 | lRsl kdKa mq31 mq41 2Adb 4Rsl 4Adb 3Rsl 2Rsl 3Adb   |
| 331 | ghz2 ghz3 ghz1 Ghazali: effondrement des philosophes  |
| 320 | Rmq1 Rmq2 Rmq3 Ibn Roshd: traité décisif  |
| 334 | rZt1 rZt2 rZt3 Ibn Roshd: commentaire d'Aristote  |
| 13  | rZt4 (livre Z de la métaphysique)   |
| 1   | isft Épître sur l'Unité et la Trinité CAH:50 axes   |

En dimension 50, comme en dimension 20, les textes de philosophie se séparent; avec la même structure; à ceci près que manquent, en dimension 50, les deux fragments {TfGh, TfHy} du "philosophe autodidacte".

Il est satisfaisant que se détache d'abord du reste la branche 347, qui comprend la presse moderne et les brèves notices introduisant, dans un recueil de l'INALCO, les "textes et poèmes choisis" (cf. TxAr).

Restent, dans la branche 35, les textes littéraires anciens et contemporains. Les classes de la partition retenue sont généralement cohérentes, se prêtant à recevoir un titre global; surtout si l'on tient compte des



subdivisions distinguées par des parenthèses au sein de la classe 329. La structure d'ensemble des (89+51+37) fragments apparaît, toutefois, plus claire si l'on ne tient compte que de 20 facteurs (cf. §3.2.1).

#### 4 Conclusion: perspectives pour la lexicologie de l'arabe

L'analyse statistique multidimensionnelle a déjà été appliquée à des données très diverses quant au domaine d'origine et au format: économie, médecine, questionnaires d'opinion, expériences de psychologie, descriptions de spécimens en zoologie ou botanique; linguistique, enfin.

La preuve la plus sûre que la structure suggérée par l'analyse ne résulte pas d'une construction aléatoire, mais a un fondement réel, réside dans la concordance entre cette structure et des informations certaines que l'analyse n'a pas explicitement prises en compte.

Tel est l'intérêt du traitement de corpus. Si, ayant dénombré, au sein de fragments de textes nombreux et divers, des formes de mots, ou des locutions, on retrouve - au moins partiellement - la répartition de ces fragments en œuvres ou en genres; on a quelque raison de prêter attention à ce que les analyses suggèrent comme subdivisions au sein d'ensembles compris sous un seul titre; et comme rapprochements entre des parties, *a priori*, nettement séparées.

L'étendue du corpus considéré ici est modeste: mais la cohérence des résultats obtenus nous incite à poursuivre la saisie et l'analyse de textes; en perfectionnant par degré le choix des formes à dénombrer; et jaugeant l'efficacité des diverses méthodes et de leurs variantes.

L'expérience nous intéresse d'autant plus que la langue arabe, dans l'immensité de sa littérature, offre comme une palette continue de langues dont

il importe de déterminer le spectre. Ainsi, le problème de la création d'un dictionnaire général pourrait-il être résolu sous une forme dont on ne rêvait même pas avant l'avènement de l'informatique. Les mots seraient expliqués non seulement par des énumérations de sens ou d'acceptions, mais par une distribution de ceux-ci au sein d'une typologie des textes. Cherchant ce qu'est un mot dans un contexte, l'utilisateur recevrait pour réponse une liste d'occurrences expliquées de ce mot; occurrences prises seulement dans des textes proches, quant au type, de celui qu'il étudie.

### Références bibliographiques

[MOTS ARABES]: "Sur l'étude des textes arabes d'après les occurrences des formes de mots", in *CAD*, Vol. XIX, n°1, pp. 65-84; (1994).

[TEXTES ARABES]: "Typologie d'un ensemble de textes arabes d'après les occurrences de formes de mots et de locutions", in *CAD*, Vol. XX, n°4, pp. 389-412; (1995).

[DISCR. CAH LING.] : "Analyse discriminante et classification ascendante hiérarchique dans l'adjonction d'individus à un corpus de référence: application à des données linguistiques"; in *CAD*, Vol. XX, n°1, pp.45-66; (1995);

*Textes et Poèmes choisis pour l'épreuve de récitation*; Vocalisation, présentation et commentaire de Ayadi CHABIR et Jean TARDY; recueil distribué par la Section d'Arabe Littéral de l'Institut National des Langues et Civilisations Orientales (INALCO) ; Centre d'Asnières; (1995).

*Documents pour l'étude et l'histoire de la langue arabe* ; 1, prose; par Ayadi CHABIR , INALCO, Section d'Arabe; (1995);

[THÈSE] : Ayadi CHABIR : *Approche lexicométrique et lexicque-index des Séances d' al Hamadānī*; Paris; (1997).