

J.-P. BENZÉCRI

F. BENZÉCRI

**Programmes de statistique linguistique : mise  
à jour et applications à l'étude du contenu  
de textes bibliques en grec**

*Les cahiers de l'analyse des données*, tome 16, n° 2 (1991),  
p. 133-160

[http://www.numdam.org/item?id=CAD\\_1991\\_\\_16\\_2\\_133\\_0](http://www.numdam.org/item?id=CAD_1991__16_2_133_0)

© Les cahiers de l'analyse des données, Dunod, 1991, 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/>

## **PROGRAMMES DE STATISTIQUE LINGUISTIQUE: MISE À JOUR ET APPLICATIONS À L'ÉTUDE DU CONTENU DE TEXTES BIBLIQUES EN GREC**

**[LING. TRI 2]**

*J.-P. & F. BENZÉCRI*

### **0 Progrès des dénombrements linguistiques**

Les programmes présentés dans [LING. TRI] (in *CAD*, Vol XV, n°1, 1990), ont déjà servi à plusieurs études, dont les dernières parues sont celles de [TEXTES GRECS] (in *CAD*, Vol XVI, n°1, 1991). Cependant ces programmes, tout en conservant la même structure quant à la création et la gestion de fichiers, ont reçu récemment d'importants perfectionnements.

D'une part, la mise en forme des textes prend en compte des particularités de certaines langues: valeur diacritique des accents en français, adjonction des lettres {ch, ll, ñ} à l'alphabet en espagnol...; d'autre part, sur le judicieux conseil d'André SALEM, nous effectuons désormais tous les tris en mémoire centrale; ce qui permet d'achever en quelques minutes des traitements qui requéraient auparavant plusieurs heures!

Le présent article rend d'abord compte de ces progrès; en entrant (au §1.1) dans des détails que nous soumettons à l'attention des informaticiens, mais que les autres lecteurs pourront passer sans rien perdre de la suite. De plus, il propose aux linguistes les résultats d'analyses prenant en compte non seulement des formes de mots outils (comme dans [TEXTES GRECS]), mais aussi d'autres mots, choisis parmi les plus fréquents; et, notamment, de nombreux mots pleins.

Selon nous, le dénombrement des mots outils (même s'il convient encore de le perfectionner) apparaît comme une méthode stylistique susceptible de fournir une typologie générale des textes d'une langue.

En prenant en compte les mots pleins, on aborde un problème d'un autre ordre de grandeur: celui de l'analyse automatique du contenu. Nous croyons qu'ici il faut considérer explicitement, non seulement le profil de distribution des mots, mais leur répétition par grappes plus ou moins nombreuses; ce qui permet de distinguer entre concepts généraux et mots pertinents représentatifs de la diversité d'un corpus: ceux-là omniprésents; ceux-ci plus étroitement resserrés,

et offrant en grappes, dans quelques textes, la plupart de leurs occurrences. (cf. A. AÏT HAMLAT, [IND. DOC.], in *CAD*, Vol IX, n°2, 1984; et [TEXT. DOC.], *ibid.*).

Cependant, sans reposer sur une méthode dont les perspectives générales nous apparaîtraient clairement, les résultats obtenus en prenant en compte des formes de mots pleins sont d'une netteté remarquable; et, ce qui est mieux, font voir d'un point de vue nouveau la composition de certains textes bibliques.

## 1 La bibliothèque des programmes

Nous ne reprenons pas l'exposé de [LING. TRI], mais considérerons successivement, la gestion du tri par fusion en mémoire centrale; les programmes de mise en forme des textes de diverses langues; les performances dans le tri et la création de concordances.

### 1.1 Tri par fusion de formes en mémoire centrale

André SALEM a adopté le tri après codage pseudo-aléatoire des formes ("hash-coding"): il lui est réservé de nous exposer comment il met en œuvre cette méthode classique. En partie parce que les incertitudes de ce codage nous rebutent, et plus encore afin de ne rien changer aux accès de notre système de programmes, nous avons conservé la méthode, suffisamment rapide et tout aussi classique, du tri par fusion.

La seule difficulté rencontrée vient de ce que les formes de mots (accompagnées de leurs adresses dans le corpus; et éventuellement d'un commentaire ou d'un contexte comme dans la création de concordances) constituent des enregistrements de longueur variable; afin de loger ces enregistrements sans aléat ni dépense rédhibitoire de mémoire, nous avons accepté d'affecter à chaque forme un pointeur; et même deux, pour le tri par fusion. Voici quelques détails, destinés, répétons le, au lecteur informaticien.

De façon précise, considérons un texte, que nous supposons saisi comme un fichier ASCII 'D:Texte'. Selon la procédure exposée dans [LING. TRI] (et que nous nous bornons à résumer ici), 'D:Texte' n'est pas soumis tel quel au programme 'trimu§' de tri par fusion d'une liste de mots, mais doit être converti en un autre fichier de texte 'D:Texte§' où chaque mot du texte initial (plus exactement chaque occurrence de forme), est écrit seul sur une ligne; avec, sur la ligne suivante, le numéro du chapitre et celui du verset (ou plus généralement de l'alinéa, suivant le numérotage introduit dans le texte initial); l'ordre initial des mots étant conservé. Éventuellement, si l'on désire créer une concordance, la deuxième ligne afférente à un mot comporte, après les deux numéros de chapitre et de verset, ce même mot inséré dans son contexte (cf. *infra* §1.3).

Le listage 'D:Texte§t', créé par le programme 'trimu§', est de même format que 'D:Texte§'; à ceci près que les formes y sont recopiées dans l'ordre alphabétique; chacune répétée autant de fois qu'elle a d'occurrences dans le texte

'D:Texte'; avec les adresses rangées dans l'ordre croissant, par chapitres et versets; (c'est-à-dire, en fait, dans leur ordre initial; que 'trimu§' respecte sans s'enquérir aucunement du contenu de la deuxième ligne qu'il recopie).

Nous verrons au §1.2 que nous disposons désormais de programmes (propres aux diverses langues) permettant de passer directement de 'D:Texte' au fichier trié 'D:Texte§', sans créer de fichier d'occurrences 'D:Texte§'. Mais quoi qu'il en soit de ce perfectionnement, nous voulons ranger en mémoire centrale le contenu structuré de 'D:Texte§'; et avoir aisément accès à tout enregistrement correspondant à une occurrence de rang déterminé (e.g. celle de rang 8973...).

```
type taw=array[0..4095]of ptr;ptaw:^taw;
      tbb=packed array[0..32767]of byte;ptbb:^tbb;
var teza,tezt:array[0..255]of ptaw;tbxx:array[0..255]of ptbb;
```

À cette fin, en tête du programme de tri sont déclarés des types: 'taw' est, en bref, une zone de mémoire de 32k destinée à ranger des pointeurs; on accède à de telles zones par des pointeurs, 'ptaw'; dont deux tableaux 'teza' et 'tezt' de 256 sont prévus, ce qui permettrait de ranger un million de pointeurs; pour accéder, par deux voies, aux enregistrements afférents à un million d'occurrences. De même, le type 'tbb' est destiné à loger ces enregistrements; et on y accède par des pointeurs 'ptbb'.

Au cours du déroulement du programme, des zones de mémoire de 32k sont succesivement demandées pour satisfaire aux besoins. Au départ, on alloue des zones aux trois pointeurs teza[0], tezt[0], tbxx[0]; et, chaque fois qu'une nouvelle occurrence se présente, on vérifie qu'on a où la loger; et, si besoin est, on initialise une nouvelle zone; soit pour les pointeurs; soit pour les enregistrements).

```
new(teza[0]);new(tezt[0]);new(tbxx[0]);
.....
if (wa=4096) then begin
  wa:=0;za:=za+1;new(teza[za]);new(tezt[za]);end;
readln(tVr,ma);readln(tVr,qa);
lgx:=length(ma)+length(qa)+1;
if (bxx+lgx>32767) then begin
  tt:=tt+1;bxx:=0;new(tbxx[tt]);end;
```

Dans le fragment de programme publié ici, 'wa' est le numéro du nouveau pointeur à demander dans la zone en service, laquelle a elle-même pour numéro 'za'; si wa = 4096, il faut allouer de nouvelles zones pour les pointeurs (par les ordres new(teza[za]) et new(tezt[za]; za ayant été augmenté de 1). De même, 'ma' et 'qa' sont les deux parties (forme proprement dite et numéros), de l'enregistrement afférent à une nouvelle occurrence; 'bxx' est le nombre d'octets déjà utilisés dans la zone en service pour loger ces enregistrements (zone numéro 'tt'); s'il ne reste pas assez de place, on alloue une nouvelle zone (par l'ordre new(tbxx[tt]); tt ayant été augmenté de 1).

```

pont := @tbxx[tt]^[bxx]; bxx := bxx + lgx + 1;
teza[za]^[wa] := pont; tezt[za]^[wa] := pont;
pons := stringptr(pont); pons^ := ma;
pons := stringptr(@pons^[length(ma) + 1]); pons^ := qa;

```

Il est essentiel que la mémoire destinée à loger les enregistrements soit allouée par grandes zones (ici: de 32k); car il serait très lent d'affecter une nouvelle zone de mémoire pour chaque enregistrement. En terme technique, on se garde de demander d'affecter ( $lgx+1$ ) octets au pointeur  $teza[za]^[wa]$ , mais on prend ces octets dans la zone  $tbxx[tt]^[bxx]$ , qui est en cours de chargement; on fait simplement pointer  $teza[za]^[wa]$  (ainsi que:  $tezt[za]^[wa]$ ) sur le premier octet non utilisé  $tbxx[tt]^[bxx]$ ; et par un jeu, propre au Pascal, de typage de pointeurs, on copie successivement, à partir de cet octet, ce qui constitue les deux lignes 'ma' et 'qa' de l'enregistrement. (En lisant le fragment de programme, on se souviendra que pour loger une chaîne de longueur  $lg$ , il faut  $lg+1$  octets, dont le premier indique précisément la longueur de la chaîne).

Quand est achevée l'installation du texte formaté en mémoire centrale, est associé à toute occurrence de rang  $nn$  ( $nn$  = rang dans le texte initial, 'D:Texte'; ou, ce qui est équivalent, rang dans 'D:Texte\$'), un enregistrement vers lequel pointent les deux pointeurs  $teza[zz]^[ww]$  et  $tezt[zz]^[ww]$ , dont les indices  $zz$  et  $ww$  sont calculés par les formules:

$$zz := nn \text{ div } 4096 \quad ; \quad ww := nn \text{ mod } 4096 \quad ;$$

les opérations de tri s'effectueront sans déplacer les enregistrements, mais en permutant seulement les valeurs des pointeurs; de telle sorte qu'en fin de tri, le même pointeur  $teza[zz]^[ww]$  se trouve pointer vers l'enregistrement qui a le rang  $nn$  dans la liste ordonnée 'D:Texte\$'.

L'algorithme de tri en mémoire centrale est formellement identique à celui donné au §1.2 de [LING. TRI]; mais au lieu de faire défiler deux copies 'Vra' et 'Vrb' d'un même fichier en cours de tri, pour recopier, sur un troisième fichier 'tVr', le premier (dans l'ordre lexicographique) des enregistrements qui se présentent sur les deux voies, on simule cette opération en comparant les enregistrements vers lesquels pointent deux pointeurs mobiles  $ptr_a = teza[za]^[wa]$  et  $ptr_b = teza[zb]^[wb]$ ; et on recopie, au rang approprié, dans 'tezt' celui des deux pointeurs,  $ptr_a$  et  $ptr_b$ , qui pointe vers l'enregistrement qui est le premier dans l'ordre lexicographique.

L'identité formelle est si parfaite que le nouveau programme 'trimu\$2', de tri en mémoire centrale, a été créé, à partir du programme 'trimu\$' de fusion de fichiers, en effectuant des modifications ponctuelles sans qu'il soit besoin de prendre garde à la gestion ordonnée des fusions partielles entre paires de segments de longueur, 1, 2, 4, 8, 16, ...; gestion, dont le principe associatif est expliqué dans [ASS. FUS. TRI] (cf. K. Ben SALEM, in *CAD*, Vol XV, n°2, pp. 133-138; 1990).

## 1.2 Mise en forme des textes de diverses langues

Pour créer à partir d'un texte, 'D:Texte' un autre fichier de texte, 'D:Texte§', où chaque mot du texte initial (plus exactement chaque occurrence de forme), est écrit seul sur une ligne; avec, sur la ligne suivante, le numéro du chapitre et celui du verset (ou plus généralement de l'alinéa, suivant le numérotage introduit dans le texte initial), nous utilisons actuellement des programmes de mise en forme spécialement adaptés aux particularités de quatre langues: respectivement 'forgalac§2', 'forhsp§2', 'forlat§2' et 'forgrec§2' pour le français, l'espagnol, le latin et le grec.

Ces programmes ont en commun plusieurs fonctions: reconnaître le découpage du texte en mots, les signes de ponctuation étant éliminés; réduire toutes les capitales en minuscules; introduire les numéros de chapitres et verset (en avertissant l'utilisateur d'une éventuelle incohérence dans le numérotage; si un numéro de verset est répété, ou qu'il manque). Ces fonctions sont expliquées dans [LING. TRI], §2.1: la seule modification apportée ici est que, même si le programme n'a pas rencontré de numéros, il attribue aux formes le nombre 1 (et non zéro) pour le chapitre et le verset.

Dans l'actuelle version '2', les programmes 'for' comprennent une première phase de lecture du texte 'D:Texte', où le fichier 'D:Texte§' des enregistrements (avec deux lignes par mot) est seulement créé en mémoire centrale (selon la disposition expliquée au §1.1); et une deuxième phase d'écriture, où 'D:Texte§' est effectivement écrit sur disque.

La séparation des deux phases fait gagner du temps; de plus, elle suggère d'assurer par un seul programme la mise en forme et le tri, sans créer effectivement sur disque de fichier 'D:Texte§' (des occurrences dans leur ordre naturel), mais en passant directement au fichier 'D:Texte§t' (fichier des occurrences rangées alphabétiquement); une phase de tri des formes, (très rapide parce qu'effectuée en mémoire centrale), étant interposée entre la lecture et l'écriture. Ont été combinés à cet effet les programmes 'trigalac§2', 'trihsp§2', 'trilat§2' et 'trigrec§2'.

Quant aux particularités des langues, il semblerait naturel de concevoir un programme paramétré prévoyant toutes les combinaisons usuelles de doubles lettres et de signes diacritiques: accents du français, esprits du grec... Mais nous ne savons pas encore tout ce qu'il convient de prendre en compte. Car, par exemple, l'accent grec dépend du contexte et son interprétation requerrait une analyse grammaticale complexe; de plus, sur la police de caractères, l'accent peut être combiné avec l'esprit: on se borne à éliminer l'accent grec pour ne garder que l'esprit. En espagnol, nous éliminons simplement l'accent; le caractère unique 'í' étant, e.g., remplacé par 'i'.

En français, il nous paraît suffisant de ne tenir compte de l'accent que s'il est susceptible de distinguer entre deux formes: comme 'chante' et 'chanté', ou

'chantes' et 'chantés'. En fait, l'accent français n'est pris en compte que s'il affecte une voyelle qui, dans le mot considéré, n'est suivie d'aucune autre voyelle, mais est la dernière lettre ou n'est suivie que d'une consonne. Ainsi, on confond le nom 'complètement' (action de compléter), avec l'adverbe 'complètement' (de façon complète): mais une telle confusion nuit peu à l'analyse statistique.

Nous avons cru devoir respecter strictement l'ordre alphabétique des mots. À cette fin, les signes diacritiques conservés sont rejetés à la fin du mot.

En français, les accents aigu, grave et circonflexe sont notés respectivement: '/', '"', et '>'. On a ainsi les couples {chante, chante/}; {chantes,chantes/}, {a, a"}, {du, du>}. Parce que, dans le code ASCII, les caractères {/, ", >} ont un ordre inférieur à celui de toutes les lettres {a, b,...,z}, les mots accentués prennent place, comme il convient, avant tout mot comptant une ou plusieurs lettres de plus: 'du>' avant 'dual'; 'a"' avant 'abri'; 'ré', codé 're/', avant 'récit', codé 'recit'.

En grec, l'esprit rude, seul conservé, est noté '<'; d'où le couple {εἰσ, εἰσ<} (préposition 'vers' et numéral 'un'); le iota souscrit (conservé seulement s'il affecte la dernière lettre d'un mot) est noté '!'; ce qui (avec une police appropriée) donne le couple {κεφαλή, κεφαλή, } ('tête', au nominatif et au datif). Ici encore, {<, !} précèdent toutes les lettres dans le code ASCII: ainsi, εἰσ< précède εἰσαγωγή; et τη, précède τησ.

Le cas de l'espagnol est particulier en ce que les dictionnaires de cette langue traitent 'ch', 'll' et 'ñ' comme des lettres uniques, distinctes des 26 autres et insérées dans l'alphabet selon l'ordre suivant:

{a, b, c, ch, d,... , l, ll, m, n, ñ, o,...} ;

Afin d'obtenir ce même effet, nous codons {ch, ll, ñ} suivant les combinaisons {cl, l{, n~}; dont le deuxième caractère a, dans le code ASCII, un ordre supérieur à celui de toutes les lettres usuelles {a, b,... , z}. Ainsi, les mots:

{cocodrilo, coche, chiste}, écrits {cocodrilo, cocle, cliste},

sont classés comme ci-dessus; alors que, laissés tels quels, les conventions alphabétiques usuelles leur donneraient l'ordre inverse:

{chiste, coche, cocodrilo}.

De même les couples {luz, llama} et {antes, año}, codés {luz, l{ama} et {antes, an~o}, se placent dans leur ordre espagnol qui est inverse de celui du français.

Le cas du latin est le plus simple: on se borne à supprimer les trémas, les accents qui peuvent avoir été introduits; et à éclater les combinaisons de lettres: {æ, œ, ñ, fl}.

### 1.3 Performances des programmes

Prenons pour exemple le texte grec du livre du prophète Isaïe, dans la version des Septante, traité sur un Macintosh SE/30. Sauf erreur de notre part dans la saisie, le texte comporte 26977 occurrences.

Soumettons ce texte 'D:Isa' au programme 'forgrc§2': la phase de lecture (qui comprend aussi la mise en forme d'enregistrements, avec deux chaînes de caractères par occurrence) dure environ 70s; la phase d'écriture 55s: au total, le fichier d'occurrences 'D:Isa§' est créé en 2 minutes environ. L'espace requis en mémoire est de 7 segments de 32k pour les pointeurs; et de 10 segments de 32k pour les enregistrements.

Soumettons 'D:Isa§' au programme 'trimu§': l'entrée prend environ 70s; la phase de tri en mémoire centrale dure 3' 40s; l'écriture de 'D:Isa§t' s'effectue en 55s (cette opération étant identique à l'écriture de 'D:Isa§'; qui comprend les mêmes enregistrements; dans un autre ordre). La mémoire requise comprend d'une part deux fois 7 blocs de 32k pour les pointeurs; et d'autre part 10 segments pour les enregistrements; soit, au total, 24 fois 32k = 768k.

En soumettant le texte initial 'D:Isa' au programme 'trigrec§2', on crée 'D:Isa§t' en un seul passage; avec trois étapes, dont la durée nous est déjà connue: 70s pour l'entrée; 3' 40s pour le tri; 55s pour l'écriture. L'espace utilisé est, comme avec 'trimu§', de 768k.

Proposons-nous de créer une concordance. Le programme 'prCcrd§2' effectue simultanément la lecture de 'D:Isa§' et la création d'un nouveau fichier qui ne diffère du précédent qu'en ce que, après l'adresse de chaque occurrence, celle-ci est répétée, sur la même ligne avec un contexte comprenant, avant et après, un même nombre de formes qu'on peut choisir de 2 à 5. Dans le cas présent, on a choisi 3 comme longueur des contextes droit et gauche. Et le temps de création a été de 4' 50s pour le fichier 'D:Isa(3§'.

Le fichier 'D:Isa(3§' offre par lui-même, peu d'intérêt: il doit être trié. On le soumet à 'trimu§2'; (en donnant, cf [LING. TRI] §3.3.2, pour nom du texte de base, le nom fictif 'D:Isa(3').

On ne s'étonnera pas que le temps de lecture soit considérable, du fait de la longueur des deuxièmes lignes des enregistrements: 5' 45s. L'occupation en mémoire est de même accrue: toujours deux fois 7 blocs de 32k pour les pointeurs; mais 45 blocs pour les enregistrements eux-mêmes; soit, au total, 59 fois 32k = 1888k.

En revanche; le temps de tri reste de 3' 40s; puisque le tri s'effectue sur pointeurs et que seule est prise en compte la première ligne de l'enregistrement; laquelle, dans 'D:Isa§' comme dans 'D:Isa(3§', est réduite à l'occurrence seule. Enfin, la sortie de la concordance proprement dite, 'D:Isa(3§t', prend 2' 10s.



καρδια

- 1 5 πονον και πασα καρδια εισ λυπην απο  
 6 10 επαχυνθη γαρ ης καρδια του λαου τουτου  
 15 5 αυτησ γνωσεται ης καρδια της μωαβιτιδος βοα,  
 19 1 αυτου και ης καρδια αυτων ηττηθησεται· εν  
 21 4 μη βλεπειν ης καρδια μου πλαναται και  
 29 13 με ης δε καρδια αυτων πορρω· απεχει  
 32 4 δωσουσι και ης καρδια των ασθενουντων προσεξει  
 32 6 λαλησει και ης καρδια αυτου ματαια νοησει  
 44 20 οτι· σποδος ης καρδια αυτων και πλανωνται  
 66 14 και χρησεται ης καρδια υμων· και τα  
 καρδια,  
 6 10 ακουσωσι και τη, καρδια, συνωσι και επιστρεψωσι  
 9 9 υβρει· και υψηλη,· καρδια, λεγοντες πληνθοι πεπτωκασιν  
 38 3 μετα αληθειασ εν καρδια, αληθινη, και τα  
 44 18 του νοησαι τη, καρδια, αυτων και ουκ  
 46 8 πεπλανημενοι επιστρεψατε τη, καρδια, και μνησθητε τα  
 47 7 ταυτα εν τη, καρδια, σου ουδε εμνησθησ  
 47 8 ης λεγουσα εν καρδια, αυτησ εγω ειμι  
 47 10 και ειπασ τη, καρδια, σου εγω ειμι  
 49 21 ερεισ εν τη, καρδια, σου τισ εγεννησε  
 51 7 μου εν τη, καρδια, υμων· μη φοβεισθε  
 57 1 ουδεις εκδεχεται τη, καρδια, και ανδρες δικαιοι  
 60 5 και εκστηση, τη, καρδια, οτι· μεταβαλει εισ  
 καρδιαν  
 40 2 λαλησατε εισ την καρδιαν ιερουσαλημ· παρακαλεσατε αυτην  
 46 12 οτι· απολωλεκότεσ την καρδιαν οτι· μακραν απο  
 57 11 ουδε εισ την καρδιαν σου και εγω  
 57 15 τοισ συντετριμμενοισ την καρδιαν ουκ εισ τον  
 61 1 τουσ συντετριμμενοισ την καρδιαν κηρυξαι αιχμαλωτοισ αψεσιν  
 65 16 αυτων επι την καρδιαν εσται γαρ ο·  
 65 17 αυτων επι την καρδιαν αλλ ευφροσυνη και  
 καρδιασ  
 44 25 και μαντειασ απο καρδιασ αποστρεφων φρονιμουσ εισ  
 59 13 και εμελετησαμεν απο καρδιασ ημων· λογουσ αδικουσ  
 63 17 σου εσκληρυνας τασ καρδιασ ημων· του μη  
 65 14 τον πονον της καρδιασ υμων· και απο

À la demande du Docteur Distel, spécialiste d'électrophysiologie et d'Écriture Sainte, nous retenons de cette concordance les occurrences des formes du mot 'cœur', qui se rencontre à divers cas (nominatif, datif, accusatif et génitif singuliers; et accusatif pluriel, de même forme que le génitif singulier; mais aisément distingué de celui-ci par le contexte: article, préposition...). Le tableau publié ici ne diffère de la partie correspondante du listage 'D:Isa(3&t)' qu'en ce qu'on a laissé subsister qu'une seule fois chacune des formes {καρδια, καρδια,·, καρδιαν, καρδιασ} répétée dans le listage brut avant chacun des contextes où elle se trouve.

Un lecteur auquel le texte des Septante est familier n'a pas de peine à évoquer, d'après le contexte court du listage, l'association dans laquelle rentre le mot cœur: souffrance, chagrin, sincérité, trouble, endurcissement... La locution *συντετριμμένοι τὴν καρδίαν*, "qui ont le cœur contrit, broyé" (littéralement, "broyés quant au cœur"), se trouve deux fois: en Is.57,15 et en Is.61,1. Comme le participe broyé n'est pas au même cas dans les deux occurrences, la répétition ne serait pas décelée par un programme se bornant à enregistrer des formes sans en faire l'analyse morphologique.

L'égarément est évoqué trois fois par des formes différentes du même verbe: *πλανᾶται* en Is.21,4; *πλανῶνται* en Is.44,20; et *πεπλανημένοι* en Is.46,8. Dans le premier cas, cœur est sujet du verbe; dans le deuxième "leur cœur est cendre et ils sont égarés"; ce verset est d'ailleurs diversement construit en hébreux, grec et latin, sans que le sens diffère; dans le troisième, on appelle les égarés à manifester dans leur cœur un mouvement de repentir, de conversion...

Ce mouvement est évoqué trois fois par des formes du même verbe (plus exactement de composés d'un même verbe simple; la même racine *ב'ו'* servant aussi dans les trois cas en hébreu): *ἐπιστρέψατε*, impératif, en Is.46,8, déjà cité; *ἀποστρέφων*, participe présent, l'agent étant Dieu, en Is.44,25; et *ἐπιστρέψωσιν*, subjonctif évoquant la malédiction de ceux qui, dans leur cœur, ne comprennent pas, et ne se convertissent pas, en Is.6,10.

## 2 Analyse des chapitres du Livre d'Isaïe

L'expérimentation sur les programmes nous a incité à poursuivre l'étude des textes en croisant les chapitres du Livre d'Isaïe avec divers ensembles de formes; le but étant de sortir de la stylistique des mots outils (traitée dans [TEXRES GRECS]) pour aborder l'analyse du contenu. De multiples analyses et classifications ont été faites d'après divers lexiques de formes. Nous ne considérerons ici que deux ensembles de formes, que nous appellerons les lexiques 'dd' et 'de'. Ces deux lexiques sont extraits de la liste des formes dont la fréquence est au moins égale à 20 (les formes du mot 'cœur', considérées au §1.3, n'en peuvent donc être). Les noms propres sont éliminés.

Dans 'dd' sont retenues les quelque 148 formes satisfaisant à ces conditions, qu'il s'agisse de mots outil ou de mots pleins, tels que noms, adjectifs, ou verbes. De 'de', on a écarté les formes de l'article, du pronom relatif et des démonstratifs; les outils retenus étant des prépositions, conjonctions, voire adverbes (*οὕτως* = ainsi); et les pronoms personnels des premières et deuxièmes personnes du singulier et du pluriel; dont les formes ont précisément été écartées des analyse de [TEXTES GRECS], parce que (ce qu'on vérifiera ici) leur emploi dépend fortement du thème des textes.

c | Partition en 8 classes : Sigles des chapitres de la classe numéro c

---

121	Is01	Is55	Is50	Is33	Is59													279++++
123	Is40	Is42	Is56	Is65	Is66	Is61	Is35	Is13	Is34	277++	281=0---							251++

---

120	Is15	Is23	Is24	Is26						284--	272++	276++++	251++	
124	Is03	Is05	Is02	Is29	Is31	Is09	Is18	Is11	Is19	Is10	Is17	Is08	Is28	Is21
	Is06	Is32	Is30	Is04	Is25	Is07	Is27	Is58	Is16	Is62	Is14	Is12	Is52	Is20
	Is22								284--	272+++				

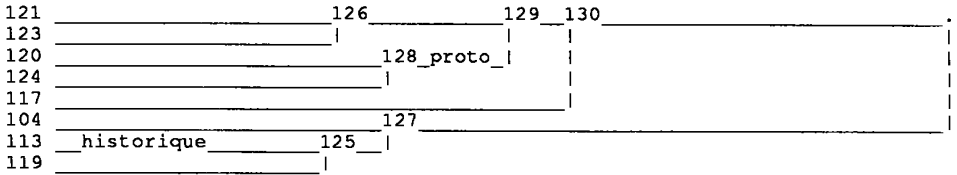
---

117	Is64	Is53	Is63	(Is63-17)						277++	282++++		
-----	------	------	------	-----------	--	--	--	--	--	-------	---------	--	--

---

104	Is54	Is47	Is60									281++++	
113	Is36	Is39	Is37	Is38	historique					267++	278++++		
119	Is44	Is45	Is51	Is57	Is49	Is46	Is41	Is48	Is43	263+	192++++	272-	288-

---



CLAS	PDS	INR	F 1	CO2	F 2	CO2	F 3	CO2	F 4	CO2	F 5	CO2	F 6	CO2
121	75	75	10	0	5	0	-528	762	151	62	-214	125	135	50
123	136	54	-56	21	215	316	-144	142	72	36	63	27	-259	458
120	46	72	-292	148	-270	127	227	90	42	3	-508	449	-325	183
124	416	75	-215	704	26	11	72	79	-3	0	44	30	108	177
117	34	107	-139	17	-863	636	-422	152	-324	89	345	102	-66	4
104	44	83	423	263	-347	176	356	185	480	338	121	21	105	16
113	73	55	154	86	29	3	225	183	-402	586	-173	109	95	33
119	176	118	478	937	80	26	4	0	-16	1	46	9	-81	27

**2.1 Analyse fondée sur un lexique de 148 formes**

On a soumis à l'analyse des correspondances le tableau (148 × 66) croisant les 148 formes du lexique 'dd' avec l'ensemble des 66 chapitres d'Isaïe. De cette analyse même, on donne seulement, sous forme d'histogramme, la distribution exacte de l'ensemble des chapitres projeté sur l'axe 1; mais on présente en détail la classification ascendante hiérarchique des chapitres et celle des formes, effectuées dans l'espace engendré par les 6 premiers axes factoriels.

On remarque d'abord, sur la CAH des chapitres, que, sur les 33 cap. que compte la classe 128 = 120 ∪ 124, trente sont numérotés de 02 à 32 (Is.13 manquant seul): la cohérence de ce bloc a été remarquée des spécialistes (ainsi que l'explique notamment le regretté G. E. WEIL, in Prat Ling, pp. 140, sqq.); le tableau des centres des classes indique que 120 est très corrélé à l'axe 1 négatif. Se signale également la classe 113 (F4 << 0), formée des quatre chapitres 36 à 39, rendant compte, en des termes qui sont ceux du Livre IV des Rois, de

γ	Partition en 15 classes : formes de la classe numéro γ
263	α< εστι< τισ< ινα< απ< αμα< μου< ει< ουτω< λεγει< ειπε< με
192	εμου< εποιησα< ειμι< εγω
278	συ< λεγω< αυτο< θεοσ< παρα< περι< ειπεν< προσ<
281	σε< βασιλεισ< σοι< σου
282	κυριε< ημασ< ημων<
279	μοι< υμων< αι< ουαι< αν< εαν< υδωρ< κυριον< υμιν< προσωπον< υμασ<
274	κυριος< εκ< οτι< παντες< γην< ο< εγενετο< μη< σπερμα< ταυτα< επι<
271	ιδου< ταισ< μετα< δε< θεου< ου
268	ην< ουκ< ουδε< αλλ< τα< αγιον< δοξα< εθνη< ηξει< ψωσ<
267	ημιν< εστιν< νυν< ονομα< απο< το< κρισιν< τασ< ωσ< ουχ< σιων< λαοσ<
265	αυτην< παν< ον< επ< ου< την< τω< αυτον< τι< αυτω< ουτωσ< τουσ<
277	τον< αλλα< εν< τοισ< η< και< δοξαν< εθνων< εσονται< εισ< υπο< οικ<
272	αυτοσ< αυτου< ανθρωποσ< οροσ< δοξησ< πασα< αιωνα< λαον< κυριου< εξ<
276	παντα< αυτοισ< αυτουσ< πνευμα< αυτων<
272	ουκετι< τροπον< κατα< εσται< ταδε< σαβαωθ< τη< ημερα< εκεινη<
251	δια< πολλοσ< οταν< εψ< τουτο< εωσ< τησ< γησ< ημερασ< επι< γαρ< η<
251	αυτη< εκει< γη< αυτησ<

la réalisation de prophéties d'Isaïe se rapportant au règne d'Ézéchiass et contenues dans les chapitres précédents.

Cependant, la classification des chapitres doit être expliquée par un étiquetage en terme de classes de formes issues de la CAH de celles-ci. Soit, e.g. la classe j117 de cap. fortement associée à la classe i282 de formes {κυριε, ημασ<, ημων<}, {Seigneur, vocatif; nous, accusatif et génitif}; l'association se voit dans cette invocation, en Is.63,17:

Τί ἐπλάνησας ἡμᾶς, κύριε, ἀπὸ τῆς ὁδοῦ σου; ἐσκλήρυνας τὰς καρδίαις ἡμῶν... (Seigneur, pourquoi nous as-tu laissés errer loin de ta voie; pourquoi avoir laissé nos cœurs s'endurcir).

Soit encore la classe j121, étiquetée i279++++; le verset Is.59,3 montre comment peuvent s'accumuler des occurrences du pronom υμων<, 'de vous' = 'vos', 'votre', forme de la classe i279:

Αἱ γὰρ χεῖρες ὑμῶν μεμολυσμένααι αἵματι, καὶ οἱ δάκτυλοι ὑμῶν ἐν ἁμαρτίαις· τὰ δὲ χεῖλη ὑμῶν ἐλάλησεν ἀνομίαν, καὶ ἡ γλῶσσα ὑμῶν ἀδικίαν μελετᾷ. (Vos mains sont souillées de sang,

263		284	287					294	
192									
278									
281								293	
282									
279					291	292			
274		285	289						
271		280							
268									
267		283	286						
265									
277									
272					290				
276		288							
251									

CLAS	PDS	INR	F 1 CO2	F 2 CO2	F 3 CO2	F 4 CO2	F 5 CO2	F 6 CO2
centres des 15 classes de la partition choisie pour les formes								
263	48	46	493 695	238 162	-182 95	-121 42	-44 6	8 0
192	13	57	1201 883	248 38	-12 0	-80 4	94 5	-337 69
278	25	41	352 205	30 1	350 203	-591 580	-84 12	-2 0
281	39	128	690 398	-583 284	367 113	451 170	69 4	194 31
282	7	95	-335 23	-1914 741	-744 112	-731 108	85 1	-275 15
279	32	61	37 2	209 63	-543 423	145 30	-400 230	420 253
274	132	15	160 632	10 2	-6 1	-44 47	-104 269	-45 49
271	52	16	161 234	-61 34	-125 140	169 257	190 325	-34 11
268	77	20	-106 117	-170 301	-234 566	40 16	6 0	7 1
267	99	18	-90 124	93 133	81 100	-185 531	2 0	85 111
265	277	20	-88 289	96 344	22 18	39 57	87 287	12 6
277	63	38	-225 227	61 17	-120 65	-93 39	305 419	-227 233
272	39	54	-484 463	52 5	372 274	94 17	116 27	329 214
276	86	50	-262 323	-167 130	127 75	41 8	-300 423	-93 40
251	11	51	-541 174	185 20	155 14	414 102	-540 174	-930 515

vos doigts sont dans les péchés; vos lèvres ont proféré l'injustice, votre langue s'exerce à l'iniquité).

Avec la classe j124 formée de chapitres de la première partie du Livre, on a noté i272+++ , renvoyant à une classe de formes qui comprend notamment ουκετι, 'ne... plus', et τροπον, 'manière'. Les versets Is.32,3 et Is.14,24 nous éclairent sur la valeur prophétique de ces formes:

Καὶ οὐκέτι ἔσονται πεποιθότες ἐπ' ἀνθρώπους,... (ils ne se fieront plus à des hommes...)

Τάδε λέγει κύριος σαβαωθ· Ὁν τρόπον εἶρηκα, οὕτως ἔσται, καὶ ὃν τρόπον βεβούλευμα, οὕτως μενεῖ,... (De la manière que j'ai dite, ainsi sera; de la manière que j'ai dite, ainsi demeurera...)

La classe j113, des chapitres historiques, est particulièrement associée à la classe de formes i278; dans cette classe, le listage d'analyse factorielle (non publié) signale {λεγων, ειπεν, προσ} comme corrélés avec l'axe (F4<0) où se

détache aussi j113; on voit, sur le tableau de correspondance, que plus de la moitié des occurrences de εἶπεν (il a dit) sont dans j113; le listage de la concordance (cf. §1.3) montre plusieurs fois répétées les locutions:

καὶ εἶπεν Ἰσαΐας (et Isaïe dit...); καὶ εἶπεν Ἐζεκίας (et Ézéchiass dit...); ce qui s'explique par le contenu des chapitres historiques.

Dans la classe i276, qui étiquette la classe j120 de chapitres, sont réunies les trois formes {εἶπε, τῆσ γῆσ} qui composent l'expression 'sur la terre'; laquelle se trouve notamment répétée, en antithèse, dans le verset Is.26,18: "...nous avons enfanté ton esprit de salut, que nous avons fait sur la terre; nous ne tomberons pas, mais tomberont tous ceux qui demeurent sur la terre"; et ailleurs encore dans Is.26.

Dans i272 sont, de même, les formes {τῆ, ἡμερα, < εκελη, } qui composent l'expression "ce jour-là" (souvent précédée de la préposition εν, 'en'), dont le pouvoir d'évocation est manifeste; et qui vient plusieurs fois dans Is.27.

Mais certaines associations peuvent être plutôt contingentes qu'essentielles. Ainsi le bref chapitre Is.15, qui constitue une prophétie contre Moab, personnifiée dans le texte des Septante comme 'la Moabite', contient 9 occurrences de l'article génitif féminin τῆσ: Is.15 s'en trouve placé au sein de j120, elle-même fortement associée à la classe i276, où est la forme τῆσ.

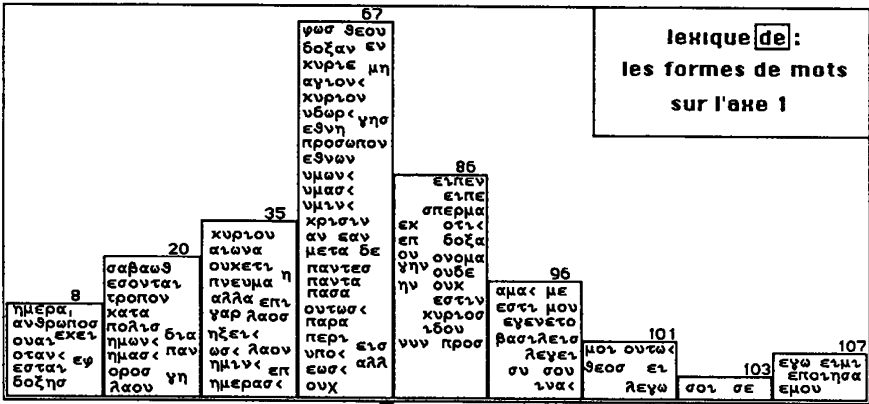
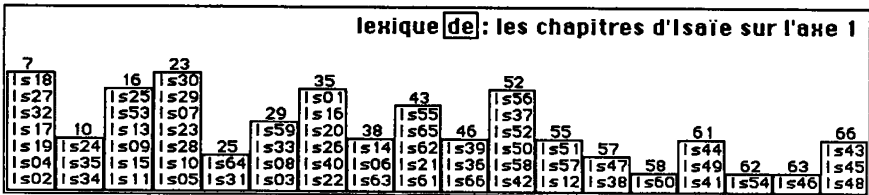
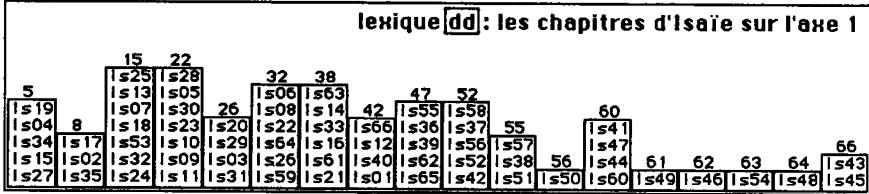
## 2.2 Analyse fondée sur un lexique de 107 formes

Les formes des mots outil les plus communs doivent-elles être prises en compte? L'article τῆσ ne semble-t-il pas avoir mal servi pour caractériser le chapitre Is.15? On a expérimenté avec un lexique 'de' de 107 formes qui diffère du lexique 'dd', utilisé au §2.1, par l'exclusion des articles, démonstratifs et relatifs. Dans leurs grandes lignes, des résultats de l'analyse du tableau (148 × 66) se retrouvent dans ceux issus du tableau (107 × 66).

148 formes de 'dd' x 66 chapitres d'Isaïe										
trace :	1.168e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	1010	686	560	507	465	431	387	363	343	324 e-4
taux :	865	587	479	434	398	369	331	311	293	277 e-4
cumul :	865	1451	1931	2365	2763	3132	3463	3774	4067	4344 e-4

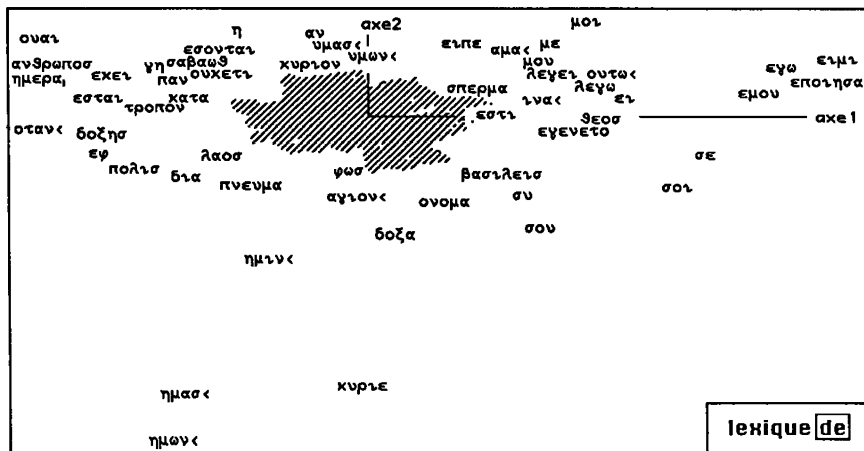
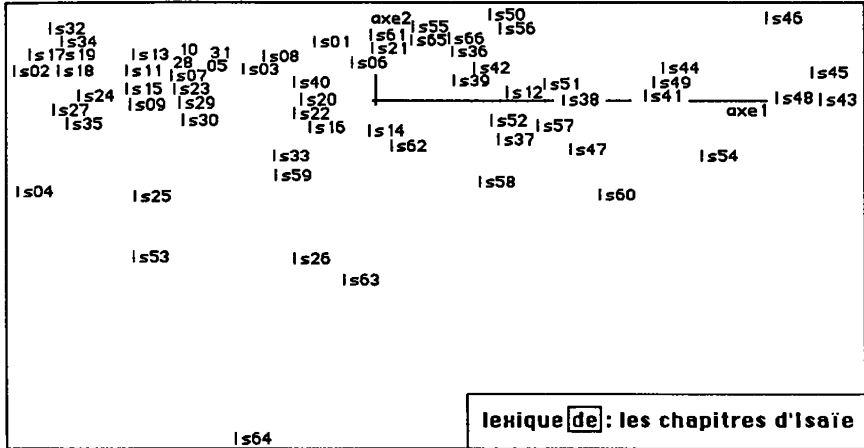
107 formes de 'de' x 66 chapitres d'Isaïe										
trace :	1.722e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	1591	1127	930	796	672	623	584	569	527	496 e-4
taux :	924	654	540	462	390	362	339	330	306	288 e-4
cumul :	924	1578	2118	2580	2971	3332	3672	4002	4308	4596 e-4

Trace et valeurs propres sont plus fortes ici que là parce que l'élimination de formes d'usage très courant, donc aux profils plats, accentue les contrastes; mais on retrouve la structure d'ensemble du livre, notamment sur l'axe 1 (sur



lequel on donne, pour les deux analyses, l'histogramme des abscisses des chapitres). Le seul avantage manifeste est qu'avec moins de formes, l'occasion s'offre de faire des graphiques lisibles. Or ces graphiques suggèrent une interprétation de la composition du Livre d'Isaïe.

Se détachent en effet, à l'extrémité ( $F1 > 0$ ) les formes nominative et génitive {εγω, εμου} du pronom 'je', avec deux formes verbales de la 1-ère personne du singulier {ειμι, εποιησα} ('je suis', 'j'ai fait'); suivies, à quelque distance, des autres formes de ce pronom {μου, με, μου}. Suit le pronom de la 2-ème personne du singulier, 'tu'; mais, pour celui-ci, les formes de l'accusatif et du datif, {σε, σοι}, s'écartent plus que {σου, συ} (génitif, nominatif). Or, dans les chapitres les plus écartés sur ( $F1 > 0$ ), la première personne qui s'affirme est Dieu; tandis que la deuxième personne est généralement celle concernée par la parole divine, que porte le prophète.



Cherchons les autres formes de pronoms personnels: {*υμασ<*, *υμων<*, *υμων<*}, accusatif, génitif, datif de ‘vous’, (le nominatif, *υμεισ<*, dont la fréquence est moindre n’étant pas compté), se projettent à l’origine sur l’axe 1. Les formes {*ημων<*, *ημασ<*, *ημων<*} de ‘nous’ sont nettement du côté (F1<0); (comme pour ‘vous’, le nominatif, *ημεισ<*, est plus rare). À l’extrémité (F1<0) le sujet n’est pas désigné par un pronom personnel, mais par le nom commun *ανθρωποσ*, ‘homme’, accompagné d’une lugubre exclamation: *ουα<*! Sont également sur (F1<0) les collectifs {*πολισ*, *λαοσ*}, {ville, peuple}. Dieu lui-même est désigné par le titre hébreux de *σαβαωθ*, traduit littéralement: ‘des armées’.



Ainsi, sur l'axe 1, du début du Livre (F1<0) à la fin de celui-ci (F1>0), selon le mode de communication entre Créateur et créature, les pronoms personnels se placent dans l'ordre {nous, vous, tu, je}; tandis que les chapitres historiques se signalent (sur F4<0) par les prises de parole d'Ézéchias et d'Isaïe lui-même.

### 3 Chapitres du Livre du prophète Isaïe et chapitres de l'Apocalypse de Saint Jean

Quant à l'emploi des formes des mots outil, l'analyse de [TEXTES GRECS] a décelé une proximité entre l'Apocalypse et le livre d'Isaïe; ceci nous incite à étudier simultanément ces deux livres en prenant en compte les mots pleins; afin de voir comment se manifeste la différence de contenu, alors que la différence de langue est relativement faible. Comme pour les chapitres d'Isaïe, de multiples analyses et classifications ont été faites d'après divers lexiques de formes.

Le lexique de base est extrait de la liste des formes dont la fréquence est au moins égale à 19 sur l'ensemble des 88 chapitres d'Isaïe et de l'Apocalypse. On a, d'emblée, éliminé les formes susceptibles d'assurer une discrimination trop facile entre les deux livres: d'une part les noms propres, d'autre part les mots pleins qui ne sont employés que dans un seul des livres.

Ainsi a été construit un tableau 'IsApC' (195 × 88) croisant un lexique de 195 formes avec les 88 chapitres. L'analyse a montré que les nombres {505εκκ, 12; ερτακ, 7} étaient fortement associées à l'Apocalypse (livre adressé à 7 Églises d'Asie Mineure; et où sont souvent mentionnées les 12 tribus d'Israël). L'analyse a donc été reprise sur le tableau 'IsApX' (193 × 88), sans ces deux nombres. Puis, comme au §2.2, on a écarté les formes de l'article, du pronom relatif et des démonstratifs; d'où un tableau 'IsApZ' (152 × 88). Quant à la clarté de l'interprétation de la CAH, l'analyse 'IsApX' nous paraît être la plus satisfaisante; mais, pour leur simplicité, nous retenons aussi des plans issus de l'analyse 'IsApZ' (qui diffère peu d'IsApX').

#### 3.1 Analyse fondée sur un lexique de 193 formes

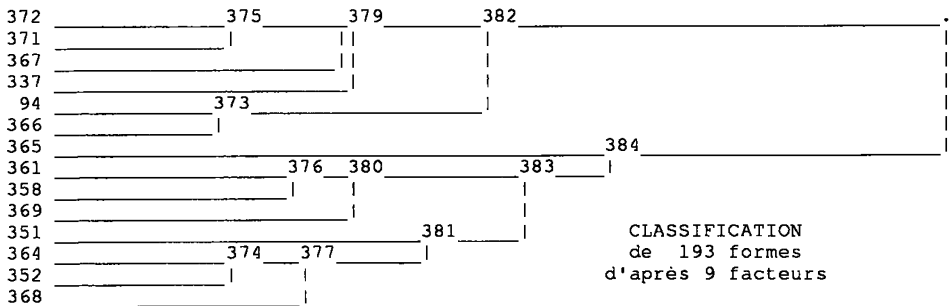
193 formes x 88 chapitres: IsApX

trace : 1.529e+0

rang :	1	2	3	4	5	6	7	8	9	10	
lambda :	1354	884	592	547	536	448	440	400	387	337	e-4
taux :	886	578	387	358	351	293	288	262	253	220	e-4
cumul :	886	1464	1851	2209	2560	2853	3140	3402	3655	3876	e-4

Nous présentons les classifications effectuées, sur les deux ensembles en correspondance, dans l'espace engendré par les 9 premiers axes; le nombre 9 ayant été choisi parce que la 10-ème valeur propre est assez bien séparée de la 9-ème. Comme au §2.1, on utilise le listage 'jVacoriq', pour étiqueter la classification des chapitres d'après celle des formes.

γ	Partition en 14 classes : IsApX : formes de la classe numéro γ
372	ακ ιδου ταισ ουχ αλλ τα ακ τασ πυρ ωσκ μετα δοξαν ην ετι εν το και της ης οταν ημερασκ επι παν θαλσσησ θαλασσαν
371	γης ης ου μη οτις ος εστιν αυτων ανθρωπων αυτοισ αυτους ουρανον τισ αυτον αυτου λεγω τω των του αυτω τουσ την
367	ουτε μετ ζωης εχων ταδε ει τι λεγει ονομα οςκ αλλα πνευμα κατα εργα δωσω
337	αυτην πολισ αυτη εκει αυτησ
94	θρονου
366	ινακ πυροσ θεου ουτοις φωνην εκ αγγελοσ εδοθη αιμακ εγενετο ουρανου μεγαλη ουρανω ηκουσα ειδον ενωπιον ειςιν θεω λεγοντεσ
365	σου βασιλεισ σοι σε
361	κρισιν εαν αν υμινκ υμασκ προσωπον υδωρκ εωσκ υμωνκ
358	βασιλεωσ βασιλευσ κυριον ειπε προσ ειπεν λογουσ
369	αγιοσκ ταυτ αυτο μοι ακουσατε ουτωκ με μου θεοσ εμου ειμι εποιησα εγω
351	κυριε ημασκ ημωνκ
364	θεον ουκ επ παρα περι ονκ πολεισ εθνη εθνων θυμου γην εις τον πασα υποκ αιωνα λαον κυριου ψυχη λαοσ γαρ γη φωνη οναι οικ εξ παντα απο
352	ουτωκ δε κυριοσ αγιονκ ψωσ ηζει δοξα συ ημινκ νυν ουτοσκ δι δικαιοσυνην ορη παντεσ ουδε ουκ υμει ακακ απ σπερμα εστι
368	προσωπου αυτοσ εφ τουτο δια η τοισ λαου δοξησ οροσ ανθρωποσ τη τροπον εσοντα σαβαωθ ουκετι διοτι εσται ημερακ εκεινηκ



On remarque d'abord, sur la classification des chapitres que, bien qu'on ait veillé à na pas favoriser la discrimination entre les deux œuvres, la dichotomie

c	Partition en 10 classes : Sigles des chapitres de la classe c : IsApX													
165	Is01	Is55	Is33	Is59	Is50	Is65	Is56	Is42	Is40	Is61	Is66	déb:2	fin:9	
164	Is30	Is04	Is32	Is28	Is08	Is21	Is06	Is12	Is22	Is07	Is20	Is19	Is17	Is18
157	Rv18	Is35	Is27	Is34	Is15	Is23	Is24							Rv18 + Is déb (29/35)
156	Is53	Is25	Is26	Is63	Is64									
158	Is51	Is49	Is48	Is43	Is46	Is41	Is44	Is45				déb:2	fin (14/26)	
163	Is16	Is62	Is57	Is52	Is14	Is58	Is47	Is54	Is60					
160	Is36	Is39	Is38	Is37										historique
126	Rv03	Rv02												
166	Rv22	Rv01	Rv19	Rv12	Rv17	Rv21	Rv09	Rv11	Rv20	Rv13	Rv15	Rv10	Rv14	Rv06
140	Rv04	Rv05	Rv07											(seul manque Rv18)
165	361+++				171_173				174					
164	368+++++			169_										
157	337+++++													
156	351+++++													
158	369+++++	167		170										
163	365+++++	352+												
160	358+++++													
126	367+++++			172										
166	371+		168	366++_										
140	94+++++		372+											

CLASSIFICATION  
des 88 chapitres  
en correspondance  
avec 193 formes

supérieure de la CAH scinde l'ensemble des 88 chapitres en 2 classes dont l'une (j172) se compose de 21 des 22 chapitres de l'Apocalypse (désignés ici par le sigle 'Rv' choisi parce que le mot grec *Αποκαλυψις* se traduit Révélation).

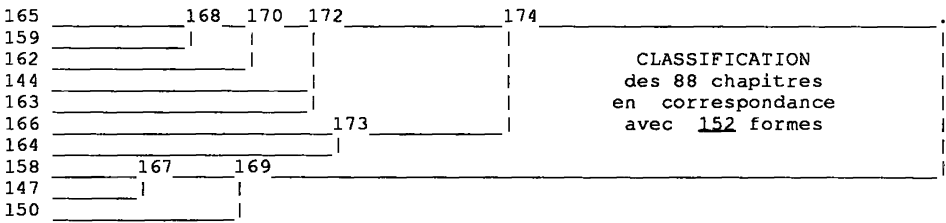
La classe j174, qui comprend tout Isaïe avec Rv18, a pu être rangé en 5 sections (dont certaines sont subdivisées), séparées sur le tableau par des lignes de tirets. La classe j160 des chapitres historique est étiquetée par i358 {βασλευσ, βασιλευσ, κυριον, ειπε, προσ, ειπεν, λογουσ} (roi, gén. et nom.; seigneur, acc.; dit; à, cf. 'dire à'; paroles, acc.); ce qui s'accorde avec le contenu de cette classe déjà vue au §2.1.

La classe j169, subdivisée en j157 et j164, comprend, outre Rv18, 29 des 35 premiers chapitres d'Isaïe; j157 est étiquetée par i337, qui contient, outre des formes de pronoms, le mot *πολις*, 'ville', remarqué au §2.2 du côté (F1<0) avec les chapitres du début du livre; j164 est étiquetée par i368, où l'on retrouve toute la classe i272 du §2.1: {ουκετι, ne...plus; εσται, sera; τη, ημερα, < εχεινη, ce jour là; τροπον, manière;...}; toutes formes dont on a noté l'association avec le début du Livre d'Isaïe.

Les classes j165 et j167 (subdivisée en j158 et j163) contiennent principalement des chapitres de la fin du Livre; j158 est étiquetée par i369 qui

c | Partition en 10 classes : Sigles des chapitres de la classe c : IsApZ

165	Is01	Is55	Is56	Is65	Is50	Is61	Is66	Is40	Is21	Is06	Rv22	6/22	déb	
159	Is46	Is42	Is51	Is12	Is44	Is41	Is49	Is45	Is48	Is43		23/27	fin	
162	Is52	Is57	Is58	Is14	Is62	Is22	Is54	Is47	Is60			+Rv22		
-----														
144	Rv03	Rv02											isolés!	
-----														
163	Is39	Is20	Is37	Is38	Is36								historique	
-----														
166	Is53	Is59	Is33	Is26	Is63	Is64								
-----														
164	Is25	Is02	Is24	Is18	Is34	Is08	Is03	Is28	Is30	Is04	Is16	Is29	Is15	Is31
	Is13	Is05	Is10	Is23	Is09	Is11	Is35	Is27	Is07	Is32	Is17	Is19	déb	26/35
-----														
158	Rv18	Rv20	Rv17	Rv21	Rv10	Rv01	Rv19	Rv11	Rv12	Rv15				19/22
147	Rv14	Rv13	Rv09	Rv08	Rv06	Rv16								
150	Rv04	Rv05	Rv07											



contient, notamment, les formes de la 1-ère personne du singulier, qu'on a vues, au §2, associées avec la fin du livre (du côté F1>0); et j163 est étiqueté par i365: le pronom 'tu'.

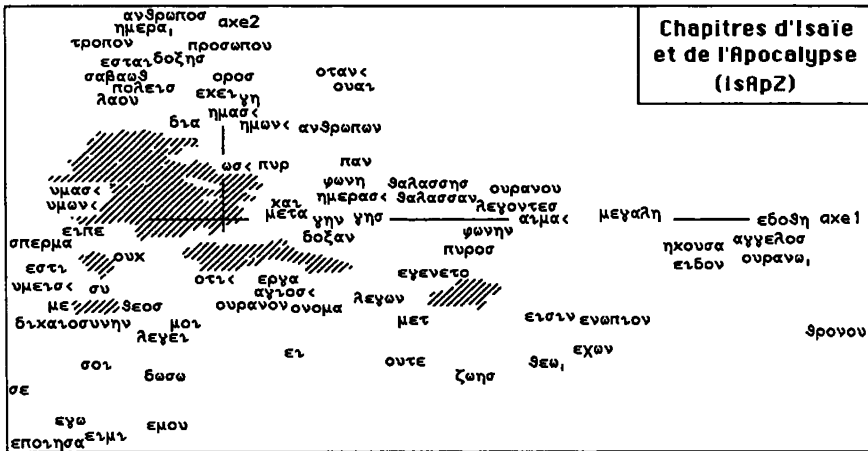
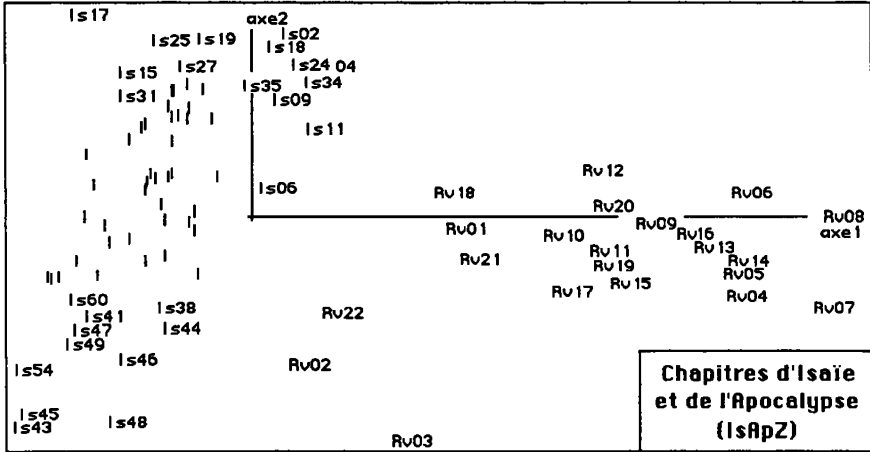
L'association de j156, {Is53, Is25, Is26, Is63, Is64}, avec i351, { $\chi\rho\rho\iota\epsilon$ ,  $\eta\mu\alpha\sigma<$ ,  $\eta\mu\omega\nu<$ } a déjà été vue au §2.2 sur le demi axe (F2<0).

La forme isolée n° 94  $\theta\rho\omega\nu\sigma$ , génitif de 'trône', qui caractérise la subdivision j140 formée de 3 chapitres de l'Apocalypse {Rv04, Rv05, Rv07}, n'est qu'une fois dans Isaïe; est respectivement {8,4,6} fois dans {Rv4,5,7}; et a 8 occurrences isolées dans l'Apocalypse.

### 3.2 Analyse fondée sur un lexique de 152 formes

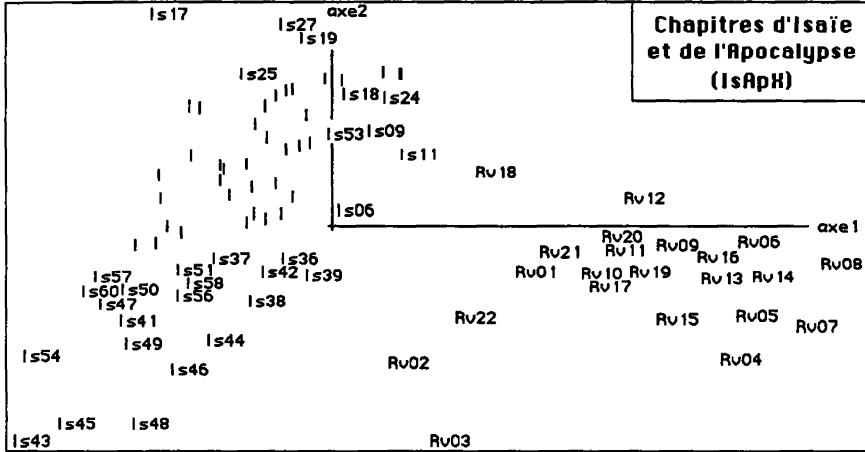
trace	:	1.839e+0												
rang	:	1	2	3	4	5	6	7	8	9	10			
lambda	:	1779	1115	762	696	641	562	550	465	437	411	e-4		
taux	:	968	607	414	379	349	306	299	253	237	224	e-4		
cumul	:	968	1574	1989	2367	2716	3022	3321	3574	3811	4035	e-4		

En écartant les formes de l'article, du pronom relatif et des démonstratifs, on obtient un tableau 'IsApZ', (152 × 88); en retenant de l'analyse de ce tableau 7 facteurs (comme le suggère la décroissance des valeurs propres), on aboutit à une CAH conforme à celle fondée sur 'IsApX', (193 × 88), mais un peu moins



claire. En revanche, il vaut la peine de considérer les plans (1,2) et (1,4), avec le nuage, assez clair, des formes de mots.

Le nuage des 88 chapitres montre d'abord (comme dans l'analyse de 'IsApX') l'étalement des chapitres de l'Apocalypse sur le demi axe (F1>0); tandis que les chapitres d'Isaïe forment une bande dense, quelque peu inclinée sur la direction de l'axe 2, et ne débordant guère du demi plan (F1<0): à la séparation des deux Livres, seuls peuvent donc faire exception, dans les CAH, les quelques chapitres de l'Apocalypse dont le facteur F1 est le moins nettement positif.

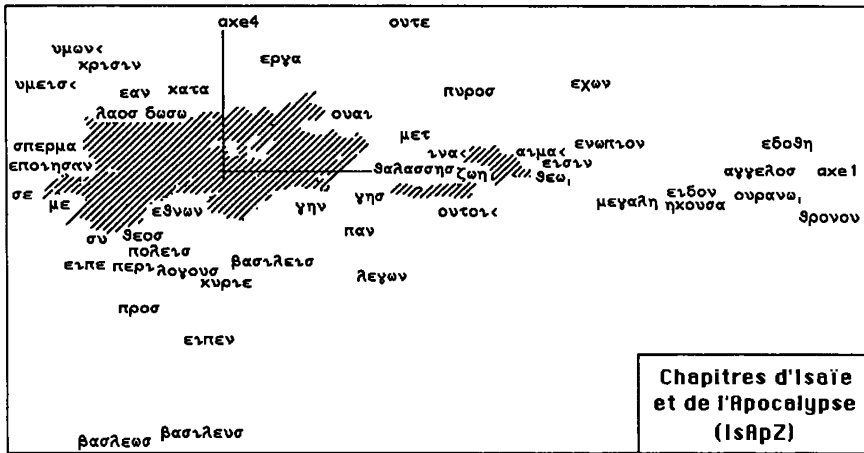
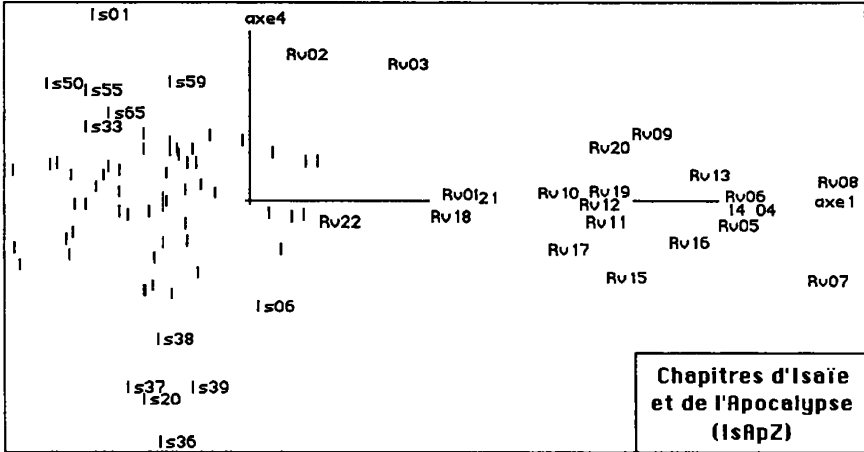


La division d'Isaïe en deux parties, initiale et finale, se déploie dans la direction de l'axe 2. Les chapitres les plus écartés du côté ( $F2 > 0$ ) sont tous numérotés entre 1 et 35; tandis que du côté ( $F2 < 0$ ) on remarque des chapitres dont le rang est au-delà de 39. Le bloc des 4 chapitres historiques {36...39}, a un facteur  $F2$  négatif, mais de faible valeur absolue: le caractère propre à ces chapitres ressortira sur le demi axe ( $F4 < 0$ ).

Le mot le plus écarté sur ( $F1 > 0$ ) est  $\theta\rho\omicron\nu\nu$ ; très proche de ce mot, dans le quadrant ( $F1 > 0; F2 < 0$ ), sont les chapitres {Rv04, Rv05, Rv07} de la classe j140: cette association a déjà été vue au § 3.1. Viennent ensuite cinq formes que nous énumérons en les accompagnant de leur traduction et de leurs fréquences respectives dans Isaïe et dans l'Apocalypse: { $\omicron\nu\rho\rho\nu\nu$ , ciel (dat.), 1+17;  $\epsilon\delta\omicron\theta\eta$ , a été donné, 2+21;  $\alpha\gamma\gamma\epsilon\lambda\omicron\sigma$ , ange, 3+11;  $\epsilon\iota\delta\omicron\nu$ , je vis (homophone de: ils virent), 7+41;  $\eta\kappa\omicron\nu\sigma\alpha$ , j'entendis (id), 5+24}.

La fréquence élevée de ces formes s'accorde avec le thème de l'Apocalypse. Alors que le génitif et l'accusatif { $\omicron\nu\rho\rho\nu\nu$ ,  $\omicron\nu\rho\rho\nu\nu$ } sont plusieurs fois dans l'Isaïe des Septante, cette version n'offre qu'un seul exemple du datif  $\omicron\nu\rho\rho\nu\nu$ , : (Is.34,5): "mon glaive s'est enivré de sang dans le ciel";  $\alpha\gamma\gamma\epsilon\lambda\omicron\sigma$  a, dans l'Apocalypse, le sens qu'a le mot ange dans la tradition des fils d'Abraham; dans Isaïe, ce sens apparaît deux fois pour le nominatif traduisant le mot hébreu  $\text{מלאך}$  (arabe:  $\text{ملاك}$ ): Is.37,36 (ange exterminateur frappant 85.000 ennemis), et Is.63,9 (ange qui sauve); mais dans Is.9,6 le mot grec est appliqué au Messie; d'autre part le sens général de messenger, se trouve, notamment dans Is.37,9: "Ézéchias envoya des ambassadeurs",  $\text{מלאכי}$ .

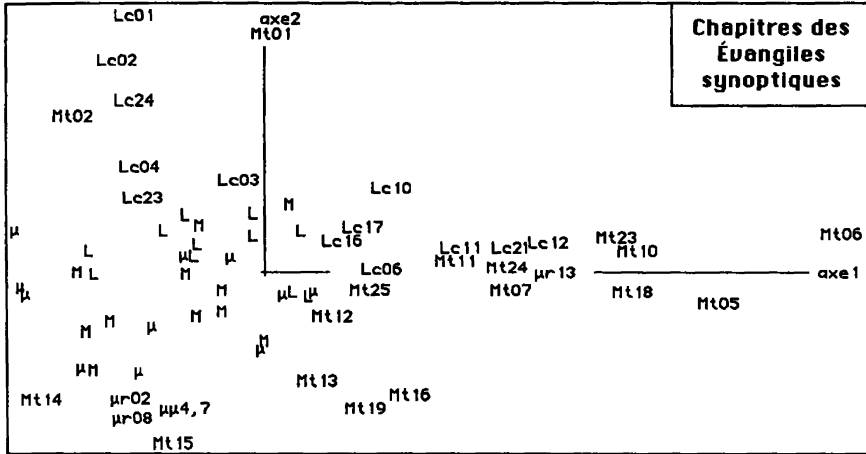
Les formes associées au début d'Isaïe ( $F2 > 0$ ), ou à la fin ( $F2 < 0; F1 < 0$ ), ont déjà été plusieurs fois considérées, dès le §2.



De même, on retrouve, du côté ( $F4 < 0$ ), les mots caractéristiques des épisodes de la vie et du règne d'Ézéchias rapportés dans les chapitres 36 à 39 du Livre du prophète Isaïe. À ces chapitres, l'analyse factorielle de 'IsApZ' et la CAH qui la suit donnent pour proche voisin Is.20; bref chapitre prophétique pour l'Égypte les Assyriens et leur roi (βασιλευσ, βασιλεωσ); et que précède une vision plus lointaine que l'actualité nous invite à citer (Is. 19,24):

Τῇ ἡμέρᾳ ἐκείνῃ ἔσται Ἰσραὴλ τρίτος ἐν τοῖς Αἰγυπτίοις, καὶ ἐν τοῖς Ἀσσυρίοις εὐλόγημένος ἐν τῇ γῆ.

ביום ההו יהיה ישׂראל שלישיה למצרים ולאשור ברכה בקרב ארצא.



#### 4 Analyse simultanée des trois Évangiles synoptiques

Tandis que le Livre d'Isaïe et l'Apocalypse de Saint Jean sont relativement proches quant à la langue mais différent quant au contenu, les Évangiles selon Saint Matthieu, Saint Marc et Saint Luc sont très proches à la fois quant à la langue (cf. [TEXTES GRECS]) et quant au contenu; de ce dernier point de vue, on a pu les diviser en un ensemble d'épisodes dont beaucoup sont rapportés dans les trois livres; ce qui permet de proposer de ceux-ci une vue parallèle ou "synopse"; d'où le qualificatif de "synoptique" attribué aux trois premiers Évangiles.

trace :	1.259e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	984	697	545	507	474	441	430	403	356	352 e-4
taux :	782	554	433	403	377	350	341	320	283	280 e-4
cumul :	782	1336	1769	2172	2549	2899	3241	3560	3843	4123 e-4

On ne doit donc pas attendre que l'analyse factorielle sépare les trois synoptiques: et il est même surprenant que le plan (1,2) présenté ici offre un nuage dont les trois pointes  $\{(F1>0), (F1<0;F2<0), (F1<0;F2>0)\}$  sont, respectivement, chargées en Matthieu (Mt), Marc ( $\mu r$ ) et Luc (Lc). Et cela, bien que, du lexique des 229 formes retenues (de fréquence  $\geq 19$ ), on ait écarté, outre les articles et démonstratifs (comme aux §§ 2.2 et 3.2), l'adverbe  $\epsilon\upsilon\theta\upsilon\sigma$  (aussitôt), dont les 48 occurrences se répartissent en 6Mt, 41 $\mu r$  et 1Lc.

Mais le résultat le plus net vient de la classification des 68 chapitres: quelle que soit la variante du lexique ou le nombre de facteurs, s'agrègent au plus bas niveau des paires ou triplets de chapitres issus de Livres différents et qui coïncident soit intégralement soit partiellement quant à la suite des épisodes qui y sont rapportés. Afin de voir avec précision comment l'inventaire des formes permet de déceler la similitude ou l'identité de contenu, on a effectué une CAH



γ	Partition en 15 classes : formes de la classe numéro γ
436	ακουσαντες δουναι οδον< παρα πατερα βασιλειαν εισελθειν αποκριθεισ μοι υιοσ εστιν ο< μεν εχων ουν ανθρωπω< καλον ουρανων
423	ολον< σωμα αρτον εμου εχετε μαθητασ λεγετε λεγειν ηρξατο εγω εδωκεν οροσ ηδη σοι οτι< εψ ει
427	εωσ< απο εκ μου ωσ< ειμι λεγων ιδου σημερον συ νυν πλην τοπον ην
428	παντες με ετι χειρασ εψη λαβων σε ποτε προσελθων κυριε
379	βασιλευσ ον< ουδε αρχιερεισ
431	τοτε ειδεν γεγραπται υπο< εκει παντασ ελθων ακουσασ κατ προφητου
402	οπισω ψυχη σημειον αρτουσ
430	ουχι βασιλεια, ουχ ημερα, < βασιλεια εσται ουτωσ< ανθρωπου εσοντα< ημερασ< δε< ημερα< ημερα<σ<
421	πνευματι πνευμα ηλθον περι παντων πασιν καθωσ εγενετο ησαν απ επι πολιν κατα αν εν
420	λεγουσιν απεστειλεν ειναι νεκρων μιαν πωσ διδασκαλε θεοσ γυναικα
433	συν ευρον< ποδα ενωπιον ωρα, κυριοσ εξ ημασ< ημιν< ημων<
418	θεον γυνη ιδων ουδε οικον προσ θεου ειπεν λαου ειπαν λεγοντες τεχνα επ
360	μητερα παιδιον
362	ονομα υιον κυριου
437	δαιμονια μητη εξεστιν υιοι ο< αναστασ μετ ηλθεν οτε< ανθρωποσ εξουσιαν εξηλθεν εισηλθεν οικιαν αλλα εισ ουκ και
414	απηλθεν υπαγε< πολλα δυνατα< εποησεν οχλοσ εξω αλλ ινα< ελεγον παλιν λεγει εισ< ερχεται μετα
442	ανθρωπον οχλοι εκειθεν ιδον μαθητα<σ οχλον μαθηται γην παραβολην δια πολλοι εχει θαλασσαν ωστε< ελεγεν γενομενησ πλοιον λογον εισιν εν<
342	γραμματεισ ουαι
432	εστε ανθρωπων υμεισ< υμων< υμασ< πασ ονοματι
435	εμε πατροσ μαλλον εμπροσθεν μηδε γησ παντα πρωτον οταν< γαρ ου ουρανου ουδε μη η
413	υμιν< αμην λεγω ενα< εαν ουρανοισ οπου< πατηρ οπωσ< σου

en tenant compte de l'ensemble des 67 facteurs; et l'on a choisi une partition des chapitres en 23 classes dont la plupart sont des agrégats de 2 ou 3 chapitres parallèles; et, simultanément, les mots ont été partagés en 21 classes dont aucune n'est d'effectif très élevé.

c   Partition en 23 classes : Sigles des chapitres de la classe numero c						
¶108	Mt01	Lc01	Lc02		362++++	nativité...
68	Lc24			433++++	(résurrection du Christ)	
2	Mt02			360++++	431+++	épiphanie
¶ 71	Mt03	Lc03		436+++...	prédication de Jean Baptiste; baptême du Christ	
	¶ NB	il est satisfaisant que j71 s'agrège à j95				
¶ 95	Lc04	µr01	Mt04		431+	début de la prédication publique du Christ
113	Lc07	Mt28	Mt08	Lc15	Lc19	Lc22 (Mt11 Lc10) Lc13 Mt20 Lc17 µr16 µr11 Mt21
¶ 92	Mt17	Lc09	µr09		428+...	transfiguration; une guérison; passion annoncée
¶ 88	Lc23	Mt27	µr15		379++++	crucifixion
25	Mt25				428++++	(paraboles, le Jugement)
¶ 69	Mt26	µr14			428++ 414++++	cène et agonie
¶ 94	Lc20	Mt22	µr12		420++++	résurrection des morts; summum mandatum; filius David
¶ 90	Lc18	µr10	Mt19		436+++...	(divorce; M, µ); enfants bénis; le jeune homme riche
¶ 82	µr08	Mt16			402++++	signe; levain; confession de Pierre; annonce de la passion et de la résurrection
¶ 84	µr07	Mt15			442++++	tradition des anciens; cananéenne
¶ 96	µr05	Lc08	437++			guérison d'un possédé; résurrection de la fille de Jaïre
¶ 81	Mt09	Lc05	µr02			paralytique guéri; vocation de Mathieu; jeûne; guérison
¶ 70	µr06	Mt14				Jean Baptiste; mult des pains; marche sur les eaux; guérison
¶112	µr04	Mt13			442++++	paraboles
110	Lc16	Lc14	Lc12	Lc11	(Mt07 Lc06)	
¶ 99	µr03	Mt12				le sabbath; main desséchée; Beelzebul
23	Mt23				342++++	"malheur à vous scribes..."
¶ 75	Mt24	µr13	Lc21		430+++	eschatologie
122	Mt10	Mt18	Mt05	Mt06		

Sur le tableau du contenu des classes, le signe ¶ désigne un agrégat immédiatement interprétable par le contenu; dont nous proposons un résumé, tout en engageant le lecteur à retourner pieusement au texte.

Parcourons le tableau de la partition des chapitres, en vérifiant une partie de l'étiquetage. Les Évangiles de la Nativité (j108 = {Mt01, Lc01, Lc02}) sont étiquetés par i362, {ονομα, υιον, κυριου} (nom, fils, du Seigneur); ces mots ont pour fréquences respectives, dans les 3 cap. de j108, {3,8,2}, {3,4,1}, {3,9,6}; les seules accumulations qu'on en trouve ailleurs sont: υιον in Mt21, 3κυριου in Mt02 (Épiphanie); chapitre caractérisé par i360 {μητερα παιδιον}, (mère, enfant), avec les fréquences respectives 4 et 7; la fréquence 4 pour μητερα se retrouvant dans Lc02 (Nativité); tandis que, dans Lc01 (Annonciation et nativité de Saint Jean Baptiste), παιδιον a la fréquence 4.

S'agrègent à un très bas niveau les chapitres de j88 {Mt27; µr15, Lc23} qui rapportent le procès du Christ devant Pilate et la crucifixion; s'y associent les mots de i379 {βασιλευσ, αρχιεπισ} (roi, pontifes): le Seigneur est accusé par les pontifes de prendre le titre de roi des Juifs.

436		449	455	456	
423	439				
427					
428					
379					
431					
402					
430	441				
421					
420	440				
418					
360	447				
362					
437	445	452			
414					
442					
342	450				
432	438				
435					
413					

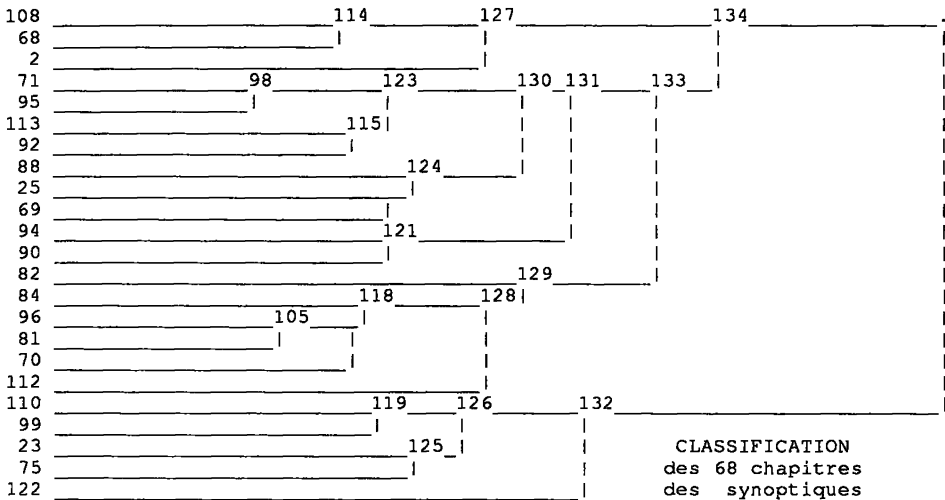
CLASSIFICATION  
de 229 formes  
d'après leurs emplois  
dans les 68 chapitres  
des Évangiles  
synoptiques.

Les chapitres de la classe j94, {Mt22, μr12, Lc20}, rapportent, notamment, tous les trois, le compte rendu d'un entretien entre Jésus et des Sadducéens qui soumettent au Maître un cas imaginé pour l'embarasser: une femme a survécu à sept frères qui ont été ses époux successifs, duquel de ceux-ci, à la résurrection des morts, sera-t-elle la femme? on trouve j94 étiqueté par i420 qui contient les formes {νεκρων, γυναικα, διδασκαλε, απεστειλεν, θεοσ, λεγουσιν,...}, (ils disent; envoya; des morts; femme, accusatif; Maître, vocatif; Dieu, nominatif;).

Voici, énumérée successivement, les fréquences d'emploi de chacun de ces mots dans les 3 chapitres, avec, éventuellement, la mention de fréquences élevées se rencontrant ailleurs: νεκρων (2,3,2; quelques autres chapitres avec 2 occurrences); γυναικα (2,4,4; qqes 2 et 3); (3,3,3); απεστειλεν (2,4,1; 4 in Mt21), employé dans la parabole de la vigne, et non à propos du cas; θεοσ (4,6,1; 4 in Lc18), en citant la parole "Je suis le Dieu d'Abraham, le Dieu d'Isaac, le Dieu de Jacob", pour en déduire: "il n'est pas le Dieu des morts mais des vivants"; (2,3,1; qqes 3), λεγουσιν n'est pas employé à propos du cas.

La classe j75, {Mt24, μr13, Lc21}, étiquetée j430+++, contient un discours qui, de l'annonce de la destruction du Temple de Jérusalem, passe à celle de persécutions, de tribulations et, finalement, de la fin du Monde. Signalons, ici encore, quelques formes caractéristiques (prises dans j430), en harmonie avec le thème du discours: futur du verbe 'être', 'jours' (cf. 'en ces jours'), 'il faut',...

εσται, (il sera = il y aura), (6,1,4; 7 in Lc01); ε̄σονται, (id, au pluriel), (2,4,2; 3 in Lc17); ημεραι, (jours, au nominatif), (3,1,2; 3 in Lc02); ημεραισ, (jours, datif), (2,2,1; 6 in Lc01, 2 in Lc02); ημερασ, (jours accusatif), (2,3,1; 3 in Lc01 et in Lc02); δει, (il faut), (1,3,1); ανθρωπου, (de



l'homme), (6,4,1; 5 *in* Mt26, 4 *in*  $\mu\text{r}14$ ), la locution “fils de l'homme” désigne le Christ, revenant pour le Jugement Dernier; βασιλεια, (royaume), (1,3,1; 6 *in* Mt13), il s'agit du Royaume des Cieux, du Ciel.

La classe j82, {Mt16,  $\mu\text{r}08$ } est marquée: i402++++; on relèra les formes:  $\alpha\text{ρ}\tau\omicron\upsilon\sigma$ , (pains, à l'accusatif), (5,6);  $\psi\upsilon\chi\eta$ , (âme), (3,3; 4 *in* Mt10);  $\sigma\eta\mu\epsilon\iota\omicron\nu$ , (signe miraculeux), (4,3; 5 *in* Lc11);  $\omicron\pi\iota\sigma\omega$ , (arrière, après), (2,2; 3 *in*  $\mu\text{r}01$ , 2 *in* Lc9), cet forme, à valeur d'adverbe et de préposition, est employée avec des connotations diverses...

En somme, la CAH a réalisé une excellente synopse; mais les formes prises en compte n'expriment pas à proprement parler le sens intégral des épisodes rapportés par les chapitres. Parfois, elles en caractérisent bien le thème; parfois la similitude nous paraît être reconnue parce que le mode d'expression est absolument identique dans plusieurs versions d'un même épisode. Ceci est particulièrement le cas pour Matthieu et Marc dont les chapitres semblent s'agrèger entre eux plus volontiers qu'avec ceux de Luc; ce qui appelle notre attention sur des agrégats de chapitres de Luc qui attestent une certaine homogénéité d'expression du troisième Évangile, plus forte parfois que l'orientation du thème.

## 5 Conclusions et perspectives

Les savants usent depuis longtemps de concordances, non seulement pour retrouver dans le texte un épisode, ou seulement une phrase, d'après un mot assez peu fréquent qui le caractérise, mais aussi pour cerner la valeur d'une forme (§1.3), en la considérant dans tous les contextes où elle se présente dans les œuvres étudiées.

L'analyse multidimensionnelle permet de considérer dans son ensemble la distribution de toutes les formes d'un lexique, choisi, ainsi qu'on l'a fait plusieurs fois ici, selon divers critères. Ainsi, au delà de succès spectaculaires dans la discrimination entre parties (§2) ou entre textes (§3), ou dans la mise en parallèle de plusieurs versions (§5), se trouve suggérée une nouvelle lecture; laquelle, peut, comme pour Isaïe, révéler une progression en termes de personnes; ou, comme pour les synoptiques, mettre l'accent sur la grande unité d'expression de plusieurs textes.

La rapidité des traitements (§1) est aujourd'hui telle que le temps à consacrer à la méditation des résultats obtenus ou à l'exploration de nouvelles voies dépasse de beaucoup celui que requièrent les opérations de calcul proprement dites.

### Références bibliographiques

[IND. DOC.]: A. AÏT HAMLAT, "Analyse des répétitions et indexation automatique des documents", *CAD*, Vol IX, n°2, pp. 173-204; (1984).

[LING. TRI]: "Programmes de statistique linguistique fondés sur le tri par fusion de fichiers de texte", *CAD*, Vol XV, n°1, pp. 59-82; (1990).

[Prat. Ling]: *Pratique de l'Analyse des Données en Linguistique et Lexicologie*; Dunod, Paris; (1981).

[TEXT. DOC.]: "Description des textes et analyse documentaire", *CAD*, Vol IX, n°2, pp. 205-211; (1984).

[TEXTES GRECS]: "Typologie de textes grecs d'après les occurrences des formes des mots outils", *CAD*, Vol XVI, n°1, pp. 61-86; (1991).

### Note complémentaire

Depuis la rédaction du présent article, les programmes ont été recompilés en utilisant TML Pascal II (au lieu de TML Pascal). Avec les nouvelles versions des programmes, les temps d'exécution sur SE/30 sont divisés environ par 4.

Reprenons l'exemple du §1.3: création d'une concordance pour le texte grec du Livre du prophète Isaïe: `forgrc§5` s'exécute en 35s; `PrCcrd§5` en 80s; et `trimu§5` en 125s; soit au total 240s, i.e. 4'.