

J.-P. BENZÉCRI

**Essai d'analyse des notes attribuées par un ensemble de sujets aux mots d'une liste**

*Les cahiers de l'analyse des données*, tome 14, n° 1 (1989), p. 73-98

[http://www.numdam.org/item?id=CAD\\_1989\\_\\_14\\_1\\_73\\_0](http://www.numdam.org/item?id=CAD_1989__14_1_73_0)

© Les cahiers de l'analyse des données, Dunod, 1989, 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/>

# **ESSAI D'ANALYSE DES NOTES ATTRIBUÉES PAR UN ENSEMBLE DE SUJETS AUX MOTS D'UNE LISTE**

**[NOTES MOTS]**

*J.-P. BENZÉCRI*

## **1 Origine et structure des données**

Monsieur Émeric Deutsch, directeur SOFRES Communication, a entrepris, selon le projet de Jean-François Steiner, de fonder les analyses d'attitudes et opinions qui lui sont demandées, sur une typologie globale établie d'après les notes données par des sujets aux mots d'une liste. Les mots proposés se comptant par centaines, et les sujets interrogés par milliers, il s'impose de recourir à la statistique multidimensionnelle. D'intéressants résultats ont déjà été obtenus, pour la SOFRES, par Mireille Gettler Summa.

Dans la présente note, nous rendons compte, brièvement, de quelques expériences que nous avons effectuées sur un tableau croisant 1140 sujets avec 286 mots notés, par chacun des sujets, sur une échelle allant de -3 à +3. Comme les données appartiennent à la SOFRES, et que la méthode et la liste même des mots, judicieusement mise au point, sont la propriété de la Société SÉMIOMÉTRICA, on ne pouvait songer à une publication sans bride.

Nous resterons donc ici au niveau général des méthodes; les résultats présentés en illustration ne seront pas complets, ni même toujours explicites, les mots étant désignés par des sigles de 4 lettres; dont certains sont suggestifs, d'autres plus obscurs; et qui ne seront développés qu'à l'occasion d'un exemple, pour nous faire comprendre du lecteur, ou pour le convaincre de la valeur des résultats obtenus.

## **2 Réponses aux questions fermées et échelles de notes**

L'introduction des méthodes numériques dans les sciences de l'opinion remonte peut-être à un siècle; et le rôle dévolu à ces méthodes croît sans cesse en importance, notamment du fait de la puissance des moyens de calcul disponibles. Sans reprendre l'exposé fait ailleurs, (cf. [QUAL. QUANT.], in

CAD, Vol XIII, n°1, 1988), nous rapellerons ici une distinction souvent faite entre qualitatif et quantitatif.

Quantité se dit, selon Aristote, de ce qui est susceptible d'égalité; et qualité, de ce qui est susceptible de ressemblance. Un ensemble de 250 notes, comprises entre -3 et +3, et attribuées par un sujet aux 250 mots d'une liste, ne constitue pas, pour la sociométrie, un objet susceptible d'égalité. Le nombre des combinaisons possibles étant extrêmement élevé, il est quasi exclu que deux sujets fournissent les mêmes notes; ou seulement qu'un même sujet, soumis une deuxième fois à l'épreuve, retrouve exactement les notes déjà données par lui. La forme chiffrée des données ne doit pas dissimuler leur caractère foncièrement qualitatif.

Une expression quantitative d'un système de notes, expression susceptible d'égalité ou, du moins, de comparaisons rigoureuses sur des échelles, ne peut être qu'au terme d'une élaboration complexe: classification, qui réduit au fini l'infini potentiel; analyse factorielle, qui, en projetant sur des espaces de faible dimension, voire sur des axes, donne à l'ensemble des possibles une densité telle qu'il se prête à la mesure.

Mais il y a plus. Les notes attribuées par les sujets aux mots ne sont pas seulement qualitatives dans leur ensemble, en tant que système complexe, elles le sont individuellement, parce ce qu'on ne sait exactement ce que mesurent ces chiffres; et qu'on est assuré que les chiffres, en tant que moyen d'expression, revêtent, pour les individus, des significations très diverses. De ce point de vue, une question fermée usuelle, admettant un petit nombre d'issues, est sans doute moins ambiguë qu'une note sur une échelle. Et, si l'on veut une échelle, un nombre minimum de degrés, (par exemple 3: {+, =, -}), a notre faveur parce qu'il laisse moins de place au jeu insaisissable des nuances.

Cependant, quand dans une même liste de mots on trouve Dieu, neige, séduire et venger, il peut sembler sacrilège de ne pas laisser au sujet la place de marquer des différences... Comment est-il seulement possible de se prêter à la consigne de noter sans réfléchir les symboles de tant et de telles réalités?

Le fait est que 90% des sujets le font d'assez bonne grâce; et que l'analyse nous convainc *a posteriori* que leur comportement n'est pas dépourvu de cohérence; qu'il manifeste, au contraire, des attitudes qui nous sont familières; suggère une typologie sur laquelle on peut espérer asseoir une sociométrie; celle, du moins, propre à la société contemporaine, à la fois normative et irrationnelle dans ses engouements et ses ostracismes. L'usage accepté des nombres algébriques atteste à lui seul que, fût-ce sans leur sens, les termes mathématiques passent dans le langage commun; et la conviction a été inculquée à tous qu'il y a une mathématique de la décision, permettant, en toute matière, de fixer par le calcul ce qui a le plus de valeur.

### 3 Correspondance entre notes et mots et entre notes et sujets

Placé devant un tableau ( $1040 \times 286$ ), l'utilisateur d'un microordinateur, même pourvu d'un disque dur et de 2 Mégaoctets de mémoire centrale, sent qu'il doit recourir à la ruse avant d'employer la force. D'ailleurs, une puissance de calcul illimitée ne suffirait peut-être pas à extraire des listes et graphiques interprétables du tableau gigantesque qu'on songe d'abord à créer, en associant à chaque mot 8 colonnes, (autant qu'il a de réponses possibles à la question afférente à ce mot), suivant le procédé désormais classique en analyse des questionnaires, (création d'un tableau sous forme disjonctive complète).

L'utilisation des notes nous paraissant une gageure, nous avons d'abord créé deux modestes tableaux à 7 colonnes:

Tableau  $S \times N$ ,  $1040 \times 7$ , avec à l'intersection de la ligne  $s$  et de la colonne  $n$  le nombre des mots auxquels le sujet  $s$  a attribué la note  $n$ ;

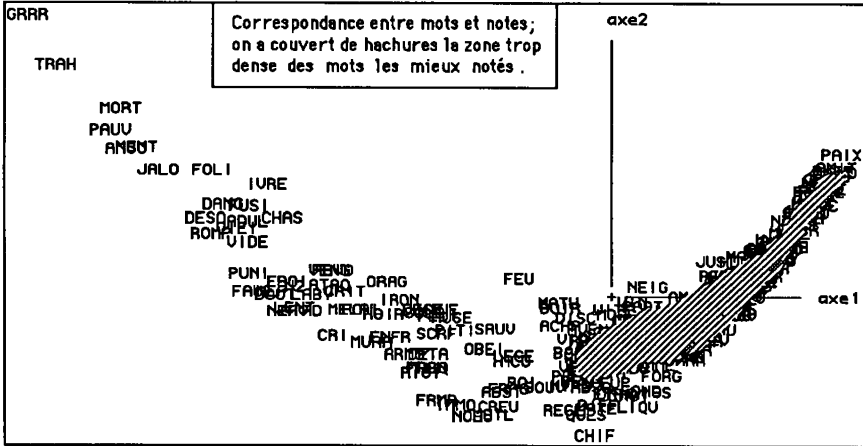
Tableau  $M \times N$ ,  $286 \times 7$ , avec à l'intersection de la ligne  $m$  et de la colonne  $n$  le nombre des sujets ayant attribué au mot  $m$  la note  $n$ .

#### 3.1 Affinités entre notes et mots

L'analyse du tableau  $M \times N$  fournit un résultat familier. Dans le plan (1,2), les notes s'ordonnent de -3 à +3, suggérant un arc que l'usage est d'appeler *parabolique*, parce que, dans le modèle classique des échelles de Guttman, le facteur 2 est effectivement une fonction de degré 2 du facteur 1. Si l'on commande que s'affichent à l'écran les sigles des mots, ceux-ci s'accumulent et s'écrasent, en dessinant un croissant, beaucoup plus dense du côté positif que du côté négatif.

Aux deux extrémités se lisent clairement les deux sigles d'une antithèse: au pôle négatif GRRR, au pôle positif, PAIX. (Signalons au passage que GRRR a été choisi de préférence à GUER, parce que le mot "guérir" figure aussi dans la longue liste, sous le sigle GRIR). Sur notre graphique, la corne de la paix, où rien ne se lisait, a été couverte de hachures; la corne de la guerre a été laissée telle quelle: on y reconnaîtra quelques mots...

Disons tout de suite que, pour naturel qu'il soit, cet étalement des mots, (depuis ceux généralement mal notés jusqu'à ceux le mieux notés), est pour nous cause d'embarras. Car d'une part, il apparaît que deux mots, dont la valeur moyenne diffère sur le marché des opinions, ne seront pas l'objet de comportements directement comparables; d'autre part, il est certain *a priori* qu'une corrélation peut exister entre des mots qui, tout en étant distants sur l'échelle moyenne, ont en commun d'être *relativement* bien notés ou mal notés simultanément par des sujets appartenant à la même tendance.



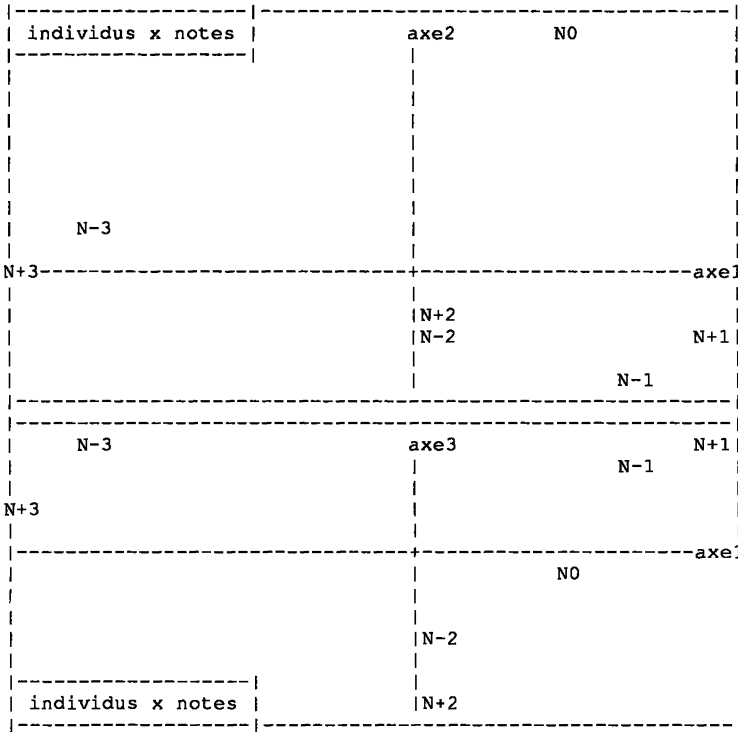
Si le croissant des mots avait une grande largeur, il y aurait, au milieu de la concavité, des cas de *conjonction des extrêmes*; c'est-à-dire des mots à la fois franchement bien notés par les uns et franchement mal notés par les autres: ces mots offriraient l'occasion de comparaisons avec tous les autres, et aideraient à appréhender un système global des valeurs. On remarque seulement, dans la concavité, les mots FEU et NEIGe; opposés à CHIFre, lequel, sur le bord convexe, apparaît recevoir de tous les sujets des notes particulièrement concordantes.

Dans un tel croissant, la classification automatique permet seulement de découper une ou deux classes d'environ cent mots, au sein de chacune desquelles les profils sur l'ensemble des notes soient assez voisins pour permettre une comparaison directe des perceptions des sujets, sans avoir à compenser, par un calcul plus ou moins arbitraire, la distance entre leurs valeurs moyennes.

### 3.2 Affinités entre notes et sujets

Tout autre est le résultat de l'analyse du tableau  $S \times N$ : il faut aller jusqu'au 4-ème axe pour trouver les notes rangées, au moins approximativement, dans leur ordre naturel. Pour surprenantes qu'elles puissent être, les dispositions des notes sur les axes précédents n'en sont pas moins classiques; la diversité des sujets quant à l'emploi des échelles numériques étant apparue, dès la thèse du regretté F. Mutumbo, dont les résultats sont publiés dans le Tome II du *Traité sur l'Analyse des Données*.

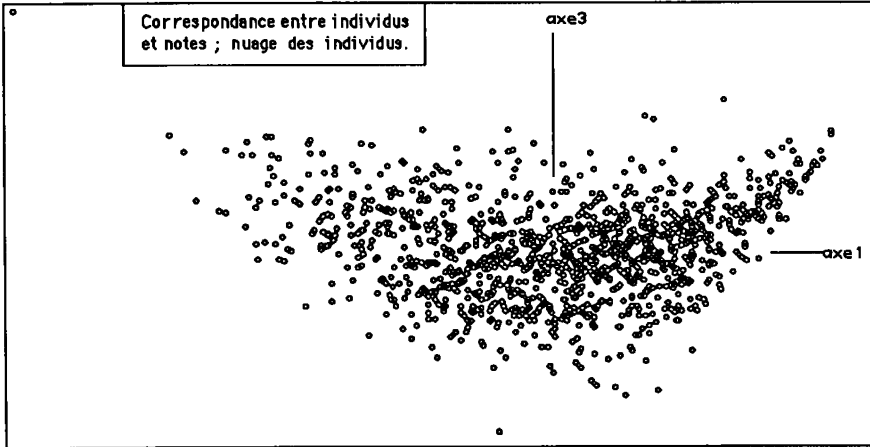
Considérons d'abord le plan (1,2). La note N0, (zéro), est très à l'écart du côté ( $F_2 > 0$ ): lui sont associés des sujets qui l'utilisent fréquemment comme



niveau de référence; ou, simplement, chaque fois qu'un mot ne leur inspire pas de réponse prononcée.

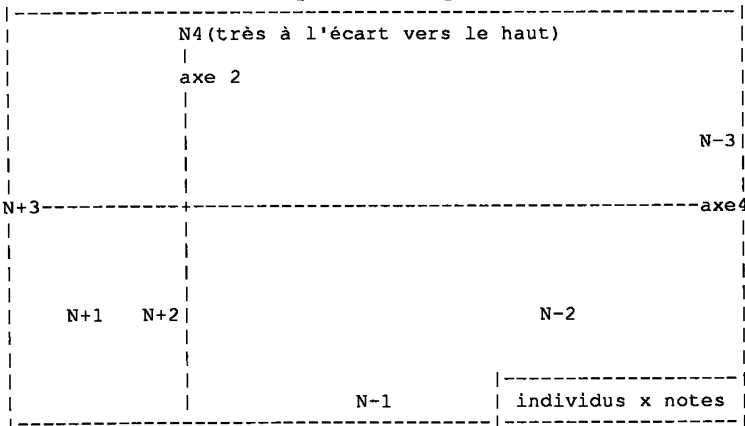
Ceci dit, les autres notes sont groupées en trois paires, rangées le long de l'axe 1 selon l'ordre des valeurs absolues. Du côté ( $F1 > 0$ ), sont les notes -1, 0 et +1, propres aux sujets qui, sans faire un usage prédominant du 0, ne s'en écartent pas souvent beaucoup. Nous disons 'pas souvent', car il est apparu que tous les sujets, à l'exception d'un seul qui n'utilise jamais le -3, recourent à l'occasion aux deux notes extrêmes; mais le font avec une fréquence très variable.

Le plan (1,3) ne nous révèle rien de nouveau quant à l'usage des notes par les sujets. On remarquera seulement que dans le nuage des notes, le zéro ne se place pas à l'extrémité du pôle de la modération (N-1, N+1). La raison en est que l'usage du zéro procède de deux tendances distinctes, sinon contradictoires. D'une part, la modération qui fait user principalement du triplet {-1, 0, +1}; d'autre part, la schématisation, qui peut conduire certains sujets à utiliser exclusivement le triplet {-3, 0, +3}.



Il est remarquable que dans le plan (1,3) le nuage des sujets revête une forme de croissant; mais ce croissant ne va pas du plus au moins; il va de la modération à l'affection pour les notes extrêmes. Un individu très à l'écart dans le quadrant ( $F1 < 0, F3 > 0$ ) a même la particularité de n'utiliser que la paire de notes extrêmes  $\{-3, +3\}$ , à l'exclusion de toute note intermédiaire.

Il faut, avons-nous dit, attendre l'axe 4 pour voir les notes de signe - s'opposer aux notes +; bien que venant après le facteur de modération, (plan



1,3), et l'attraction du zéro, (axe 2), cette opposition est sensible au niveau des sujets pour lesquels la moyenne des notes n'est pas toujours la même, mais varie; principalement entre 0 et +1.5, comme on peut le voir sur le tableau des valeurs ordonnées des moyennes, tableau créé par le programme 'zrang'.

tableau ordonne des valeurs de la variable moy									
rng	0	140	280	420	560	700	840	980	1120
7	4.01e+0	4.48e+0	4.63e+0	4.76e+0	4.86e+0	4.95e+0	5.08e+0	5.23e+0	5.60e+0
14	4.08e+0	4.49e+0	4.64e+0	4.76e+0	4.86e+0	4.95e+0	5.08e+0	5.23e+0	5.70e+0
21	4.16e+0	4.50e+0	4.65e+0	4.77e+0	4.87e+0	4.96e+0	5.09e+0	5.25e+0	
28	4.21e+0	4.50e+0	4.65e+0	4.77e+0	4.87e+0	4.96e+0	5.09e+0	5.26e+0	
35	4.24e+0	4.51e+0	4.66e+0	4.78e+0	4.88e+0	4.97e+0	5.10e+0	5.27e+0	
42	4.29e+0	4.52e+0	4.66e+0	4.78e+0	4.88e+0	4.98e+0	5.11e+0	5.28e+0	
49	4.30e+0	4.53e+0	4.67e+0	4.78e+0	4.89e+0	4.99e+0	5.11e+0	5.30e+0	
56	4.33e+0	4.54e+0	4.68e+0	4.79e+0	4.89e+0	4.99e+0	5.12e+0	5.32e+0	
63	4.36e+0	4.55e+0	4.69e+0	4.80e+0	4.90e+0	5.00e+0	5.13e+0	5.33e+0	
70	4.39e+0	4.56e+0	4.69e+0	4.80e+0	4.90e+0	5.00e+0	5.14e+0	5.35e+0	
77	4.40e+0	4.56e+0	4.70e+0	4.80e+0	4.90e+0	5.01e+0	5.15e+0	5.35e+0	
84	4.41e+0	4.57e+0	4.71e+0	4.81e+0	4.91e+0	5.02e+0	5.16e+0	5.36e+0	
91	4.42e+0	4.58e+0	4.71e+0	4.81e+0	4.91e+0	5.02e+0	5.16e+0	5.38e+0	
98	4.42e+0	4.59e+0	4.72e+0	4.82e+0	4.92e+0	5.03e+0	5.17e+0	5.40e+0	
105	4.43e+0	4.59e+0	4.72e+0	4.82e+0	4.92e+0	5.03e+0	5.17e+0	5.42e+0	
112	4.44e+0	4.60e+0	4.73e+0	4.83e+0	4.92e+0	5.05e+0	5.18e+0	5.43e+0	
119	4.44e+0	4.61e+0	4.73e+0	4.83e+0	4.93e+0	5.05e+0	5.19e+0	5.45e+0	
126	4.45e+0	4.62e+0	4.74e+0	4.84e+0	4.93e+0	5.06e+0	5.20e+0	5.48e+0	
133	4.46e+0	4.62e+0	4.74e+0	4.85e+0	4.94e+0	5.07e+0	5.21e+0	5.50e+0	
140	4.47e+0	4.63e+0	4.75e+0	4.85e+0	4.94e+0	5.07e+0	5.22e+0	5.54e+0	

pour finir l'affichage du tableau, entrer une lettre

De façon précise, on lit sur ce tableau les moyennes des notes, (décalées de 4, comme elles le sont sur le fichier qui nous a été remis), pour tous les rangs de 7 en 7. La plus faible valeur est 4.01, (i.e. 0.01), qui correspond au rang 7; la plus forte 5.70, (1.70), au rang 1234, (i.e. 1120+14). Le Maximum absolu, (rang 1140), est de 6.05, (2.05); et le minimum, (rang 1), est 2.95, (-1.05).

Que la moyenne d'un ensemble de notes entières allant de -3 à +3 soit, sur notre échantillon de sujets, largement étalée entre 0 et 1.5, est un fait qu'on doit tenter d'interpréter. Il est peu vraisemblable que soit principalement en cause l'attrance, plus ou moins forte, que les mots de la liste exerceraient sur les sujets. Une pareille explication s'imposerait si l'on soumettait à des sujets de toutes tendances une liste de noms d'hommes politiques liés à un seul parti, ou seulement une liste de noms communs ou d'adjectifs tous marqués par une tendance idéologique déterminée: dans notre cas, elle n'a pas de place.

Il nous paraît plus juste de répéter que les chiffres, en tant que moyen d'expression, n'ont pas la même valeur pour tous les sujets; sans oublier que des sujets différents peuvent vouloir donner à des sentiments ou à des opinions d'une même force des expressions soit modérées soit outrées; et que certains estiment que l'absence de louange est, sauf cas exceptionnel, la seule forme permise de critique.

Reprenant une locution d'abord introduite par les astronomes, nous dirons que les notes ou valeurs données par les sujets sont à corriger suivant l'équation personnelle de chacun. Au §5, nous tenterons de corriger les notes en les recodant suivant une formule simple et déjà plusieurs fois employée. Mais, auparavant, au §4, nous montrerons tout le parti qu'on peut tirer de l'analyse de



données univoques: nous dénombrerons, pour chaque paire de mots, les sujets ayant donné aux deux mots la même note.

#### 4 Analyse des coïncidences de notes

##### 4.1 Le tableau de correspondance

Nous dirons qu'il y a coïncidence de note entre deux mots,  $m$  et  $m'$ , quand un sujet  $s$  donne à  $m$  et  $m'$  la même note, quelle que soit celle-ci. D'où un tableau de correspondance  $M \times M$ ,  $286 \times 286$ , où  $k(m, m')$  est le nombre des sujets ayant donné à  $m$  et  $m'$  des notes qui coïncident.

Selon cette définition, on doit inscrire dans toute case de la diagonale de ce tableau le nombre 1140, effectif de l'échantillon des sujets considérés. En effet, dans la mesure où chacun des sujets n'est interrogé qu'une fois, (ce que l'on peut regretter), l'unique note donnée à tout mot  $m$  coïncide avec elle-même! Cependant, il nous a paru bon de corriger la diagonale, en donnant pour valeur à  $k(m, m)$  le maximum atteint par  $k(m, m')$  pour  $m' \neq m$ .

Ainsi, on évite que chaque mot ne tende à créer par lui-même un facteur; ou du moins que les valeurs propres, même celles de rang très élevé ne restent supérieures à un seuil. Le cas modèle d'un tableau carré symétrique, combiné linéairement avec le produit de ses marges et un tableau diagonal ayant même marge que lui, est étudié dans le traité sur L'Analyse des Données. Dans notre cas, la structure n'est pas si régulière, et l'expérience des analyses a montré que l'interprétation était plus claire si l'on corrigeait la diagonale.

##### 4.2 Premières analyses globales

On a commencé par analyser le tableau  $286 \times 286$ , sans en corriger la diagonale. Sur Macintosh +, cette analyse ne demande pas moins de 36 heures de calcul! Les premiers axes étant dominés par quelques mots, qui, même s'ils diffèrent par le sens, s'agglutinent parce qu'ils sont soit quasi universellement approuvés soit rejetés, il est difficile d'interpréter les facteurs; et nous avons recouru à la classification ascendante hiérarchique (CAH).

Aux niveaux inférieurs de la hiérarchie, sont apparues de nombreuses classes très clairement interprétables; au premier rang desquelles nous citerons:

{AME, FOI, DIEU, SACRÉ, PRÊTRE, PRIER},

classe souvent retrouvée dans la suite, telle quelle, ou peu modifiée. Ou encore:

{AVENTurier, DEFI, BOHEme, NUIT, MYSTère}.

Mais dans la CAH, comme à l'analyse factorielle, l'essentiel de l'inertie est liée à la constitution de classes de mots très souvent en coïncidence, pour la seule raison qu'ils reçoivent généralement la note maxima +3, (ou la note minima -3), sans pour autant évoquer des valeurs ou des menaces semblables.

Corrélativement, de grandes classes de mots s'agrègent à un niveau très bas; et les dernières ramifications sont confuses; d'autant plus que, la CAH étant faite dans l'espace engendré par les 29 premiers axes, il reste encore une part notable et significative de l'inertie non prise en compte, si comme dans cette première analyse, on garde la diagonale constante et égale à 1140.

Signalons ici qu'une autre analyse globale a été faite, en prenant en compte non seulement les coïncidences de notes mais aussi les inégalités de sens donné, lesquelles sont, comme les coïncidences, des données interprétables univoquement sans recodage. De façon précise, on a associé à chaque mot  $m$  trois caractères relatifs,  $m>$ ,  $m=$  et  $m<$ , avec pour tout mot  $m'$ :

$$\begin{aligned} k(m>, m') &= \text{nombre des sujets ayant préféré } m \text{ à } m'; \\ k(m=, m') &= \text{nombre des sujets ayant noté } m \text{ et } m' \text{ en coïncidence}; \\ k(m<, m') &= \text{nombre des sujets ayant préféré } m' \text{ à } m; \end{aligned}$$

d'où un tableau à 858 lignes et 286 colonnes.

Effectuée sur un Macintosh II, avec le précieux concours de la société Europe Ordinateurs, l'analyse de ce tableau a été faite en 7 heures, (avec le même programme que sur Macintosh +, donc sans utiliser adéquatement le coprocesseur arithmétique; ce qui permettrait de réduire encore grandement le temps requis).

La prise en compte des inégalités est cause que, dans cette analyse, les différences de niveau moyen entre mots, (différences clairement montrées au §3.1), ont une part encore plus grande que dans l'analyse des seules coïncidences et ne laissent pas voir nettement les différences de sens ou de tendance qui nous intéressent le plus.

### 4.3 Un exemple d'analyse partielle

Un moyen simple de s'affranchir, sans biais de codage, des différences de niveau, est de se borner à analyser et classer un sous-ensemble de mots qui diffèrent peu quant au niveau. Un tel sous-ensemble a été constitué de 93 mots qui s'agrègent à un bas niveau dans la CAH fondée sur le tableau considéré au §3.1 et ont donc, sur l'ensemble des 7 notes, des profils voisins.

Comme nous avons, d'autre part, écrasé la diagonale, ainsi qu'on l'a expliqué ci-dessus, les résultats de l'analyse factorielle et ceux de la CAH faite ensuite sont pleinement satisfaisants. Le lecteur admettra, d'après le tableau et l'arbre que nous publions, que l'interprétation des classes se fait aisément, non en terme de niveau, de pouvoir d'attraction ou de répulsion des mots, mais selon leurs sens et les attitudes qu'ils évoquent.

Il faut cependant déplorer qu'il ne reste plus que le tiers de l'intéressant vocabulaire judicieusement ourdi par J.-F. Steiner. La classe du sacré est absente; ARGEnt manque, à côté d'OR et de BIJOU; nous verrons que la classe

```

=====
  c | Partition en 14 classes: Sigles des individus de la classe numero c
=====
161| ASTU MAIT VOLO AUDA CERT ROBU FORT
153| ACCR CNCR BATI CNSR UTIL
152| CHOI ADMI DELI SOUP RAFF
148| PATI PARD CNSL NOUR PROM PROT SOIG
164| TRAD ANCT PROD TRAV ECOL REFL ENSG
168| LOGI PRCI PERF RAIS MODS SERI CHRI PRUD VRTU
158| ECRI HIST THEA POES
145| CREA INVE CHER SCIE
169| OUVÉ NOUV JEU ORIG EVAS PEAU FLEV SOMM VAGU
149| NUDI CHRN VOLU
172| MONT ESPA OCEA ILE ENSB NID FECO MOEL CHAU ANML NAGE BLEU VERT
166| ELIT HERO PUIS GLOI VIRI AMBI CONQ
171| MINC CHAT BRIL PREC ACHE MODE SUBL ADOR GRAT
167| PROP HERI OR BIJO
=====

161 _____ 175 _____ 183 _____ // _____
153 _____ | _____ | _____ | _____ |
152 _____ 173_178 _____ 181_ |
148 _____ | _____ | _____ |
164 _____ 174_ | _____ |
168 _____ | _____ |
158 _____ 176 _____ |
145 _____ | _____ |
169 _____ 179 _____ 182 _____ 184 _____ // _____
149 _____ | _____ | _____ |
172 _____ | _____ |
166 _____ 180 _____ | _____ |
171 _____ 177 _____ | _____ |
167 _____ | _____ |

```

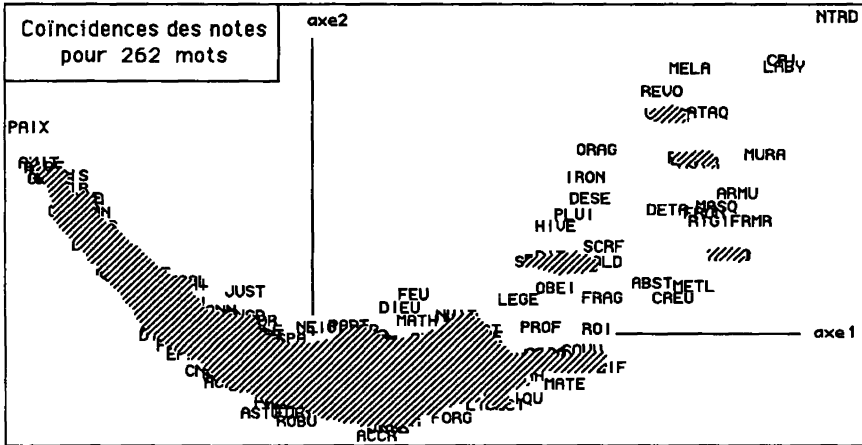
ci dessus l arbre de la partition en 11 classes

149 peut aisément s'étendre... ; ART aurait sa place, dans 158, avec THEAtre et POESie.

#### 4.4 Analyse des coïncidences pour 262 mots

Même s'il vaut la peine de constituer une liste réduite, (où les mots, ceci dit en passant, revêtiront peut-être des significations décalées, la pression du contexte étant modifiée), il conviendrait de fonder une typologie des sujets sur leurs réponses à la quasi-totalité des mots de la liste. En effet, ainsi qu'on le verra au §5, le mérite de la méthode est précisément de faire voir des tendances, (quant au style de vie ou à l'idéologie), qui, tout en étant publiquement reconnues et exprimées, sont le fait de petits groupes sensibles seulement à certains vocables.

Nous disons 'quasi-totalité', et non 'totalité' des mots. Car, par exemple, puisque le mot généralement rejeté de GueRRé reçoit le plus souvent les notes -3 ou -2, la valeur de ce mot comme pôle répulsif pour le pacifisme ou l'objection de conscience n'aura pas la place de se manifester sur notre échelle; tandis que le



rang :	1	2	3	4	5	6	7	8	
lambda :	848	82	37	15	7	4	3	2	e-4
taux :	8271	796	364	143	68	39	28	21	e-4

choix de GRRR, au nom d'une idéologie de la violence, étant présentement le fait d'une étroite minorité, sera peu clairement affirmé même sur un échantillon de 1140 sujets.

Non sans hésitation, d'après les résultats de l'analyse globale objet du §4.1, et d'après la CAH sur Mots × Notes, (tableau 286 × 7), utilisée au §4.3 pour définir une classe de 93 mots, nous avons retenu une classe de 262 mots qui seront seuls conservés dans la suite de l'étude. Ainsi qu'on l'a dit, la liste des mots étant la propriété de la SOFRES, ne peut être publiée en clair; mais les sigles suffisent pour apprécier notre sélection.

L'analyse factorielle du tableau de correspondance 262 × 262, recensant les coïncidences de notes, mais avec la diagonale écrasée, fournit des résultats beaucoup plus clairs que ceux de l'analyse globale du §4.1. L'influence du niveau général des mots est très forte, mais elle semble ne marquer que les deux premiers facteurs, où le classique nuage en croissant parabolique signe un net effet Guttman.

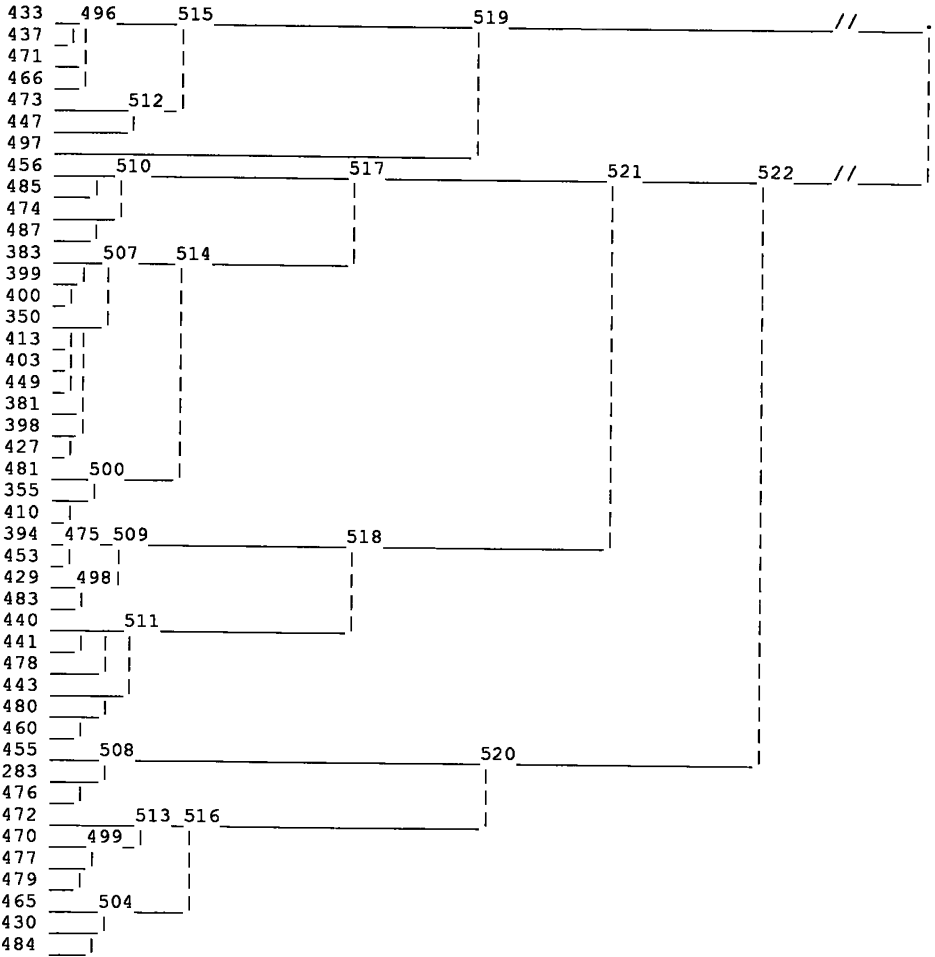
Si, à l'aide du programme 'planF', on affiche à l'écran divers plans croisant deux axes de rang supérieur à 2, on voit se détacher plusieurs îlots ou groupes de mots nettement interprétables. Bien que les facteurs de rang 1 et 2 apportent plus de 90% de l'inertie, il apparaît légitime d'effectuer une classification des mots d'après les facteurs de rang supérieur à 2; ce qu'il est facile de faire en ôtant les deux premières colonnes du tableau 'iFac' des facteurs, à l'aide du programme 'soustab'.

```

=====
c | Partition en 44 classes : Sigles des individus de la classe numero c
=====
433| SOLE RIRE GAIT TEND AMIT GRIR NAIS CNCT PEUP
437| INFI ETER
471| PAIX ABSO QUES UNIQ REGL FOND MATE CREU
466| FORG PROF LISS DIFF ABST ETRA LIQU
473| CHIF NOEU FRMR IMMO METL
447| SOUV REIN ROI
497| SACR AME PRTR PRIE FOI DIEU
456| FAMI FIDE CNFI INTE EGAL FRAN HONN
485| FRON ARMU RIGI MURA SOLD
474| MARI NOBL HERO VICT ELIT PERF GLOI
487| RECU ANCT JUST SERM INDU LOI MORA PATR HONN VRTU
383| FLER CAMP MUSI LUMI ARBR
399| IMME ESPA SOMM
400| ENSB PEAU LUNE BLEU ROUG VERT
350| VEND CMRC
413| MASQ MAIS GRAN ANCI MATE
403| CHAT CADE GRAT
449| NOIR ADOR SUBL MAGI
381| FRAG SECR HASA
398| FEMI DOUC FECO INTI
427| NTRG CARE HUMO
481| CNCR CERT BATI CHER CREA INVE SCIE OUVE IMAG
355| ENFA NID CHNT
410| ECRI HIST LIVR ART THEA POES
394| CONF RECO HERI PROP
453| RICH ARGE OR BIJO
429| COMD PUIS VIRI CONQ
483| VILL ACCR RAFF ADMI CHOI MODE PARF MINC PREC BRIL ELEG ACHE
440| AMBI VITE DEPE
441| NEIG FEU VIN BOIR
478| SAUV DESE LEGE AVEN INCO BOHE MYST EMOT CHNG ORIG
443| DESI SEDU NUDI CHRN VOLU SENS
480| NAGE ANML EAU FLEV MONT OCEA ILE NUIT VAGU
460| DEFI EVAS PART NOUV ESCA MOEL REVE SOUP CHAU JEU
455| ORAG HIVE PLUI
283| ATAQ VENG
476| MELA IRON CRIT REVO
472| DISC OBEI CERE SERV RANG ECON SCRF PITI
470| COUR PURE RESP POLI PARD CHRI
477| CNSR UTIL HUMB RAIS TRAD ECOL MATH MODR FRMT SERI PRUD TRAV PRCI PROD
479| PROT CNSL ATCH PROM DETA NOUR DELI LOGI MODS AIDE PATI SOIG ENSG REFL
465| LABY CRI FOUL NTRD
430| FORT ROBU PROG CNST AGIR EFFI DYNA
484| RUSE ENFR VOLO MAIT ASTU AUDA EFFO LUTT ACHA
=====

```

En termes géométriques, ceci revient à projeter le nuage des mots sur le supplémentaire orthogonal au plan (1,2). Une analogie simple est celle d'un objet tridimensionnel transparent allongé que l'on regarde longitudinalement; ou encore, en projection sur un plan perpendiculaire à l'axe. Ainsi, toutes les coupes transversales se superposent, et d'éventuels filets longitudinaux sont vus comme une petite tache. Ce procédé d'observation n'est toutefois pas d'une sûreté parfaite; car il se peut que l'objet allongé sinue autour de son axe.



ci dessus l arbre de la partition en 44 classes (\*mmc\*>2)

On conçoit donc que les résultats de la classification, très satisfaisants dans leur ensemble, offrent matière à critique. Afin de ne conserver que des classes bien interprétables, nous avons examiné l'arbre de la CAH générale; pour retenir finalement une partition en 44 classes, spécifiée à l'aide de 16 nœuds, comme le permet le programme 'CAH2'. Bien que J.-F. Steiner ait contesté la place de INDU, MASQ, NTRG et DETA, cette partition, qu'on peut perfectionner dans le détail à l'aide des outils logiciels dont nous disposons, fournira, après codage des réponses suivant l'équation personnelle, une typologie très suggestive.

## 5 Analyse après codage des notes suivant l'équation personnelle des sujets

### 5.1 L'équation personnelle

Le codage de chaque note  $j$  suivant trois modalités  $\{j+, j=, j-\}$ , avec une formule, dite *équation personnelle*, propre à chaque sujet, a déjà servi dans plusieurs études, (notamment [ERGO. RÉGLAGÉS] et [POLIT. GREC], in CAD, Vol XIII, n°2, 1988). En bref, on recadre entre -1 et +1 l'ensemble des notes attribuées par un même sujet  $i$ , en calculant la moyenne le Max et le min; toute note  $>$ moy est divisée par (Max-moy); toute note  $<$ moy est divisée par (min-moy). Soit maintenant  $k(i,j)$  une note recadrée; on la code sur 3 modalités par la formule:

si  $(k(i,j) \leq 0)$  alors début  $k(i,j+)=0$ ;  $k(i,j)=1+k(i,j)$ ;  $k(i,j-)= -k(i,j)$  fin  
sinon début  $k(i,j+)=k(i,j)$ ;  $k(i,j)=1-k(i,j)$ ;  $k(i,j-)=0$  fin;

Dans le cas présent, appliquer immédiatement le codage sur trois modalités à 262 variables-mot, implique de créer un tableau à 786 colonnes dont nous ne pourrions rien faire. On crée d'abord un tableau à 262 colonnes des notes recadrées; puis, à partir de celui-ci, un tableau à 132 colonnes, donnant, pour chacune des 44 classes  $q$  de la partition retenue au §4.4, les cumuls  $q+$ ,  $q=$  et  $q-$  des notes  $j+$ ,  $j=$  et  $j-$  afférentes aux mots compris dans cette classe, (cf. §5.3).

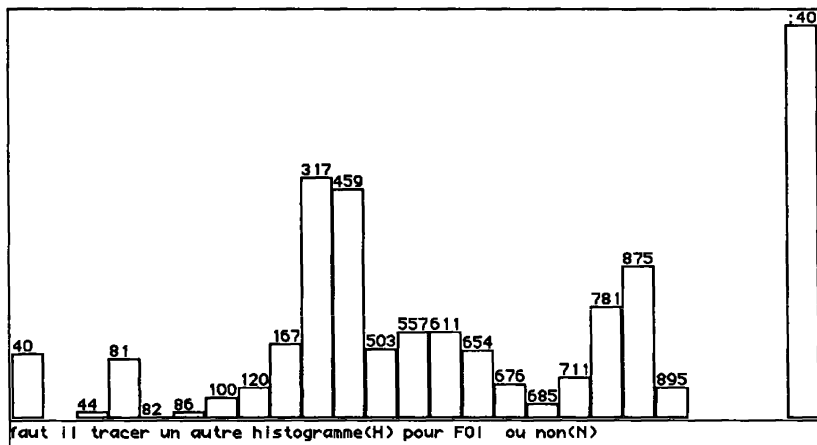
Le calcul des tableaux est rapide, pourvu que l'on insère en tête de boucles tous les calculs qu'on peut éviter de faire plusieurs fois. Dans la procédure

```

procedure burtli;begin
for j:=1 to 7 do fdj[j]:=0;
for m:=1 to 286 do begin
  jc:=jdm[m];fdj[jc]:=fdj[jc]+1;mjr[jc,fdj[jc]]:=m end;
jmi:=1;while (fdj[jmi]=0) do jmi:=jmi+1;
jmx:=7;while (fdj[jmx]=0) do jmx:=jmx-1;
jmyr:=0;tmin:=0;tmax:=0;
for j:=1 to 7 do jmyr:=jmyr+(j*fdj[j]);jmyr:=jmyr/286;
cmi[287]^i:=jmyr;
for j:=1 to 7 do begin res:=0;
  if (jmyr<j) then begin
    res:=(j-jmyr)/(jmx-jmyr);tmax:=tmax+(res*fdj[j]) end;
  if (j<jmyr) then begin
    res:=(j-jmyr)/(jmyr-jmi);tmin:=tmin-(res*fdj[j]) end;
  for r:=1 to fdj[j] do cmi[mjr[j,r]]^i:=res end;
tmin:=tmin/286;tmax:=tmax/286;tmoy:=1-(tmin+tmax);
cmi[288]^i:=tmin;cmi[289]^i:=tmax;cmi[290]^i:=tmoy;end;

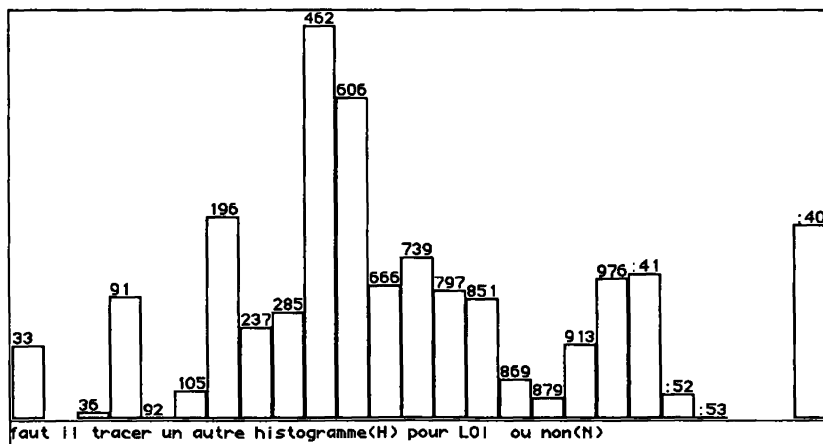
```

'burtli', on commence par dresser pour chaque sujet 7 listes contenant chacune les numéros des mots auxquels il a affecté chacune des 7 notes possibles de -3 à +3; le calcul de la moyenne est alors immédiat; le recadrage est fait successivement pour les 7 listes de mots. De plus, on calcule les moyennes, tmax, tmoy et tmin, des valeurs des modalités  $j+$ ,  $j=$  et  $j-$ , afférentes à ce sujet; moyennes qui sont un bon indice de la dispersion dissymétrique des notes initiales autour de leur moyenne.



### 5.2 Examen des notes recadrées

Il vaut la peine d'ouvrir le programme 'zrang' pour considérer, sans imprimer de tableaux, la distribution, sur l'ensemble des sujets, des notes recadrées de quelques mots, ainsi que celle des paramètres tmin et tmoyn, (la moyenne des notes brutes ayant déjà été vue au §3.2). On lit aisément les histogrammes pour FOI et LOI: le premier et le dernier créneau correspondent, respectivement, aux notes -3 et +3, avant recadrage. Sur chaque créneau est noté le rang de l'individu réalisant le maximum pour ce créneau: on sait ainsi que le premier créneau de FOI comprend 40 notes; et celui de LOI, 33; ensuite, après un créneau vide, on a un petit créneau de 4 pour FOI et de 3 pour LOI; etc... Il





ValSup	-5.21e-1	-1.38e-2	4.93e-1	1.00e+0				
	-1.00e+0	-2.68e-1	2.37e-1	6.61e-1				
-1.00e+0	7	1	4	9	4	1	3	4
-5.41e-1	3	6	5	23	9	3	6	9
-2.96e-1	5	13	14	44	26	15	21	32
-5.09e-2	10	14	32	148	45	18	58	63
1.93e-1	7	7	13	64	28	9	28	36
4.37e-1	1	1	12	11	5	17	1	18
6.82e-1	1	2	7	34	19	8	27	42
1.00e+0	6	2	10	11	5	6	6	41

tri croisant FOI (col) et LOI (lignes)

faut seulement prendre garde qu'au-delà de 999 le chiffre des centaines est remplacé par "." pour 1000, et par ";" pour 1100, selon l'ordre de numérotage des signes après les chiffres. Donc, 245 sujets ont donné à FOI le maximum +3; contre 87 à LOI. Maximum mis à part, il y a pour les deux mots une accumulation de notes inférieures à la moyenne, (laquelle, on l'a dit, varie, pour la majorité des sujets, de 0 à 1.5).

Le tri croisé, (fait suivant une maille plus large que celle des histogrammes publiés ici), montre entre FOI et LOI une certaine corrélation positive; toutefois, tandis qu'environ 50% des sujets qui mettent LOI à +3 y mettent également FOI, 40% des sujets qui mettent FOI au max, mettent LOI en dessous de la moyenne. La CAH place FOI et LOI dans deux branches majeures distinctes: on pourrait évoquer ici maint passage du Nouveau Testament...

Quant à  $t_{min}$  et  $t_{moy}$ , (respectivement taux d'utilisation par les sujets des modalités notées  $j$ - et  $j$ =, selon le codage par l'équation personnelle), il est quasi nécessaire que le tri croisé montre entre elles une nette corrélation négative; l'usage des modalités extrêmes se faisant au dépens des modalités moyennes. Cependant le jeu de compensation est complexe; et quand nous parlons de corrélation il ne faut aucunement évoquer un modèle normal.

ValSup	1.18e-1	1.49e-1	1.79e-1	2.09e-1	2.39e-1	2.70e-1	6.75e-1
	1.02e-1	1.33e-1	1.64e-1	1.94e-1	2.24e-1	2.55e-1	2.85e-1
2.20e-1	0	0	0	0	0	0	3
2.71e-1	0	0	0	1	0	1	3
3.12e-1	0	1	1	0	3	5	12
3.54e-1	0	0	0	0	4	9	9
3.96e-1	0	1	0	2	7	10	18
4.39e-1	0	0	0	5	16	18	28
4.82e-1	0	0	2	10	20	25	39
5.23e-1	0	0	0	2	30	43	47
5.66e-1	0	0	4	21	39	45	31
6.08e-1	1	0	12	31	43	33	6
6.50e-1	1	4	13	24	19	13	1
6.84e-1	1	7	16	21	4	1	0
7.35e-1	1	12	6	1	0	0	0
7.92e-1	1	1	0	0	0	0	0

pour quitter le tri croisant  $t_{min}$  (col) et  $t_{moy}$  (lignes) entrer \*

### 5.3 Cumul par classes des notes codées sur trois modalités et introduction du signallement des sujets

Puisqu'on ne peut analyser un tableau ayant 786 colonnes, on cumulera l'information. La partition en 44 classes présentée au §4.4 est acceptable, à la fois quant à l'homogénéité des classes et quant à leur nombre: c'est donc sur cette base qu'on effectue les cumulés: *a posteriori*, les résultats obtenus justifieront le procédé, tout en laissant place à des perfectionnements.

Ainsi qu'on l'a dit, à chaque classe  $q$  de mots sont associées, pour chaque sujet  $i$ , trois notes de cumul,  $q+$ ,  $q=$  et  $q-$ , suivant les formules:

$$k(i, q+) = \sum \{k(i, j+) \mid j \in q\} ;$$

$$k(i, q=) = \sum \{k(i, j=) \mid j \in q\} ;$$

$$k(i, q-) = \sum \{k(i, j-) \mid j \in q\} .$$

Si les mots d'une classe donnée recevaient tous du sujet  $i$  la même note initiale, (de -3 à +3), l'une des 3 notes  $k(i, q+)$ ,  $k(i, q=)$ ,  $k(i, q-)$  serait nulle, comme c'est le cas pour  $\{j+, j=, j-\}$ ; mais il n'en est rien, non seulement parce que l'homogénéité de la classe  $q$  n'est pas pleinement satisfaisante, mais encore parce que notre but était précisément de constituer des classes groupant des mots de niveau général différent, exerçant sur les mêmes sujets une attraction ou une répulsion relative de même force.

On notera que, pour tout sujet, le total des trois notes  $\{q+, q=, q-\}$  n'est autre que  $cardq$ , nombre des mots de la classe  $q$ ; et pour le tableau cumulé, comme pour le tableau des notes  $\{j+, j=, j-\}$ , le total d'une ligne est 262. Bien qu'on ne soit pas en format disjonctif complet, on peut donc adjoindre des colonnes supplémentaires en  $(0, 1)$ ; ou, mieux, un tableau de colonnes supplémentaires sous forme disjonctive complète.

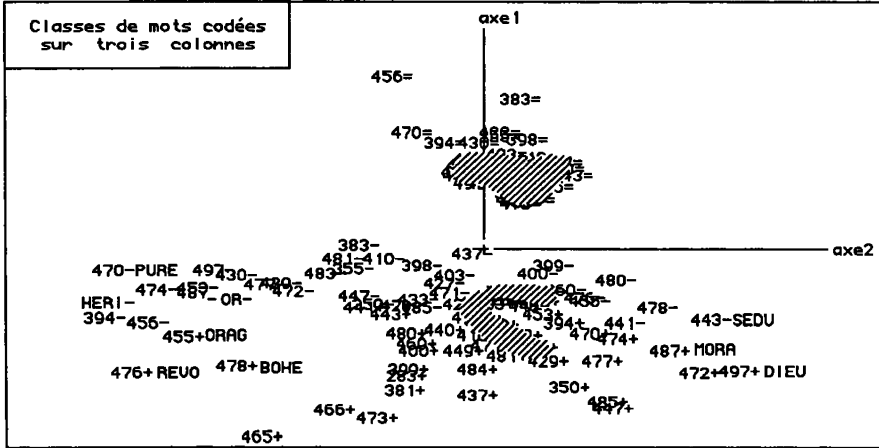
Des informations complémentaires qui nous ont été communiquées sur les sujets, nous avons deux variables:

$$\text{âge-sexe} = \{H01, H02, \dots, H10, H11, F01, F02, \dots, F10, F11\};$$

$$\text{vote} = \{LAJ, LAG, BOU, JUQ, MIT, WEC, BAR, CHR, PEN, nul, abs, nrp\}.$$

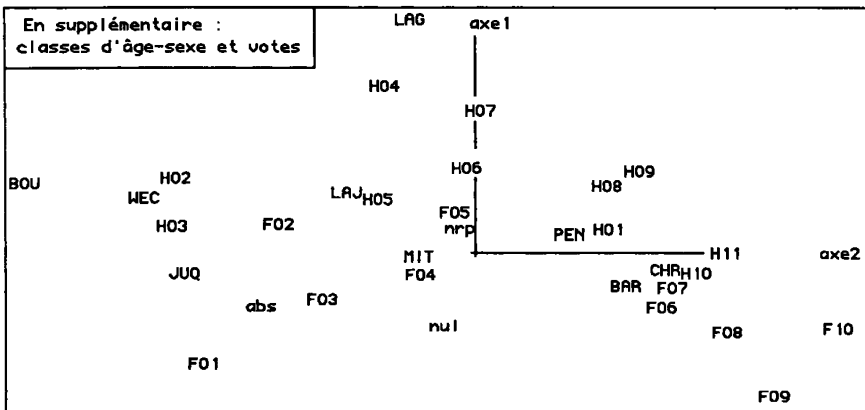
Il y a 10 tranches d'âge de 15 à 65 ans; et une au delà. Les abréviations des attitudes de votes au 1-er tour des élections présidentielles de Mai 1988 sont claires; les sujets ayant refusé de répondre sont notés 'nrp'. Au total, 34 modalités; mais en fait, il n'y a que trois garçons dans H01, et la galanterie des enquêteurs a laissé vide la classe F11...

On sait qu'au lieu de colonnes supplémentaires, on peut créer des lignes supplémentaires, centres de gravité de classes d'individus; les facteurs obtenus ici sont ceux obtenus là, multipliés par  $\sqrt{(\lambda)}$ , donc plus proches de 0.



**5.4 Résultats de l'analyse des notes recodées et cumulées**

Nous présenterons, sur des graphiques séparés, d'une part l'ensemble des variables principales, {q+,q=, q-}; et d'autre part les variables supplémentaires. Là où le nuage est trop dense, sont tracées des hachures. Pour chaque plan, l'échelle est la même sur les deux graphiques; mais la proportion entre les deux axes n'est pas respectée. Pour le contenu précis des classes, le lecteur devra se reporter au §4.4; mais nous avons placé quelques sigles de mots de ces classes. Quant aux individus, on a seulement montré, dans le plan (2,3), les sous-nuages des électeurs de Chirac et de Lajoinie, en donnant, pour chaque sujet, non son sigle, mais un caractère unique, ainsi que le permet une option du programme 'planF', pourvu qu'on ait constitué des sous-tableaux de facteurs afférents aux classes d'individus choisies.



Le graphique, confirmé par le listage des résultats montre que, sur l'axe 1, ( $\tau_1 = 19.2\%$ ), les modalités (q=), situées du côté  $F1 > 0$ , s'opposent aux modalités q+ et q-, situées du côté  $F1 < 0$ . Il s'agit donc d'une opposition entre notes extrêmes et notes moyennes, analogue à celle vue au §3.2. Il apparaît que les notes centrales sont plutôt affectées des hommes que des femmes; et que, de tous les candidats, A. Laguiller est celle dont les partisans déclarés font, (est-ce modération, est-ce désillusion?), le moindre usage des notes extrêmes.

À l'extrémité positive de l'axe 2, on a la modalité 497+: on se souvient que la classe 497 contient 6 mots du domaine spirituel; et nous avons écrit DIEU. Suivent 443-, 472+, 487+; avec 443- est rejetée la séduction etc... ; avec 487+ sont reçues la MORALE, la LOI, la VERTU.

Du côté ( $F2 < 0$ ), on trouve, le plus à l'écart, les modalités opposées à celles relevées du côté ( $F2 > 0$ ); mais également le rejet de 394, où est HERIter; et de 453, où est OR.

Il est facile de donner un nom au facteur 2; et chacun le fera bien, sans notre aide, à sa guise. Nous nous bornerons à attirer l'attention sur les modalités supplémentaires. Les candidats de droite sont, sur l'axe horizontal 2, à droite; et ceux de gauche, à gauche! Compte tenu de ce que F. Mitterrand a rallié le plus de suffrages, on ne s'étonnera pas de sa proximité au centre. Des sondages, publiés notamment par le quotidien *La Croix*, donnaient J. Chirac comme le candidat des catholiques pratiquants: ce qui s'accorde avec sa place sur l'axe 2.

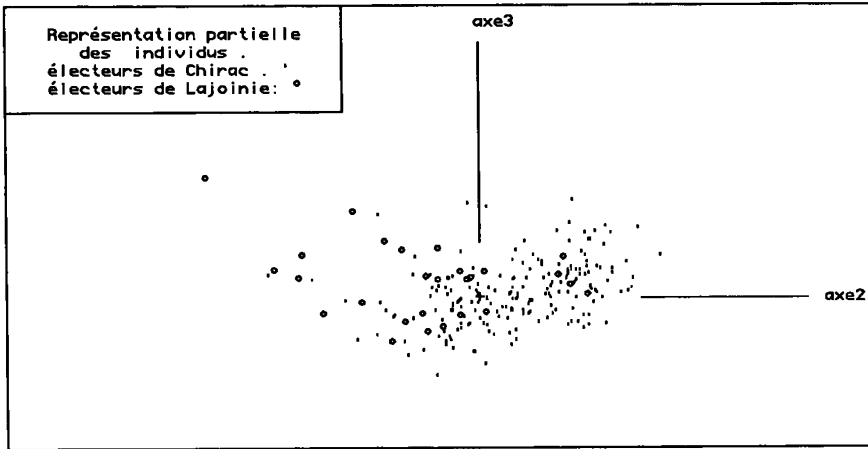
Il faut souligner que J.-M. Le Pen, n'occupe de position extrême sur aucun des 4 premiers axes. Comme F. Mitterrand, encore que par d'autres voies, J.-M. Le Pen a, en Mai 1988, rallié des électeurs fort divers; d'un tel électorat ne peut se dégager aucune vision du monde bien affirmée.

Quant aux candidats de gauche autres que F. Mitterrand, chaque lecteur verra, s'ils sont là où il les attendait!

La place des classes d'âge-sexe apparaît d'abord conforme à ce qu'on répète souvent: que la religion et la morale sont prêchées à la jeunesse par les aînés; et plutôt par les femmes que par les hommes. Seule la modalité H01, des jeunes gens, contredit à la règle; mais, comme nous l'avons dit, cette modalité ne compte que 3 individus, que nous retrouverons, par la suite, en flèche sur plusieurs axes.

Il faut ici se garder d'identifier religion, droite politique et âge mûr; il y a seulement des corrélations que l'analyse montre clairement, mais qui sont assez faibles, on le sait bien, pour être compatibles avec une grande variété de cas individuels.

Cette variété apparaît bien si l'on considère des sous-nuages d'individus ayant déclaré avoir voté pour un même candidat. Sur l'axe 2, les électeurs de J.



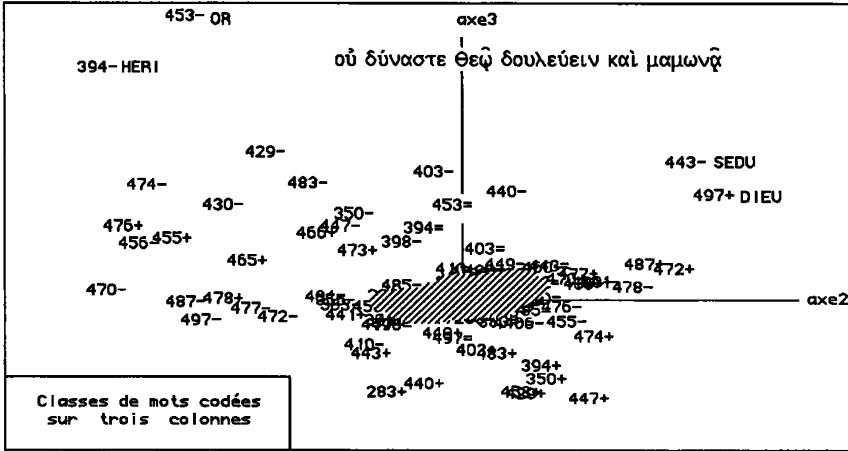
Chirac, candidat sensé être celui des catholiques pratiquants, sont, dans l'ensemble, plus à droite que ceux d'A. Lajoinie, candidat du parti communiste français: mais les deux sous-nuages ne sont aucunement séparés; ils empiètent largement, au contraire.

Nous voici transportés dans le plan (2,3):  $\tau_2 = 13\%$ ;  $\tau_3 = 6.1\%$ . Modalités des classes de mots, modalités supplémentaires et individus, semblent approximativement disposés sur un croissant; ce qui incite à dire que le facteur 3 est lié au facteur 2 par un effet Guttman.

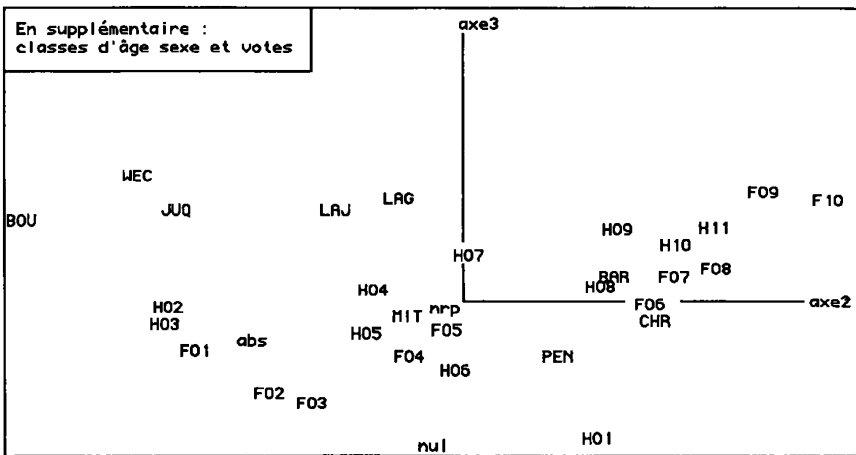
Sans rejeter cette interprétation, il vaut la peine de considérer explicitement les modalités les plus écartées sur le demi-axe ( $F_3 > 0$ ). Prises ensemble, ces modalités réaliseraient une partie du programme du *Sermon sur la montagne*; ce que rappellent, sur le graphique, quelques mots d'un verset de l'Évangile, (Matth., VI, 24). Mais le fait est qu'aucune modalité, principale ou supplémentaire, ne se projette à l'intérieur de la concavité du croissant...

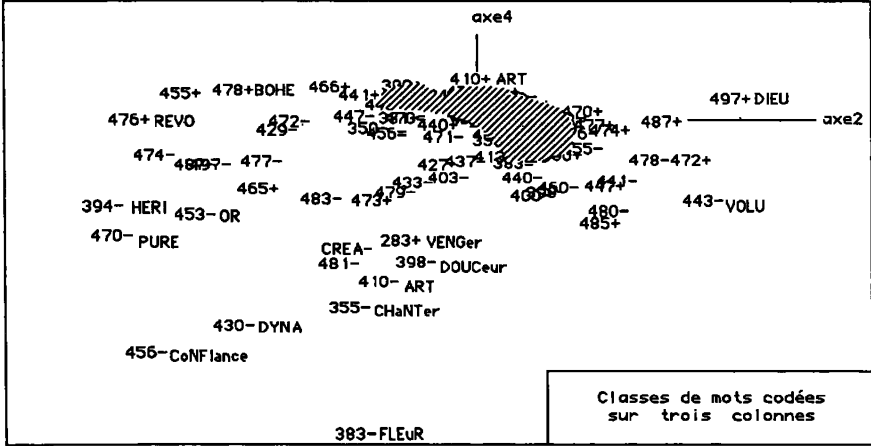
En approchant toutefois 403- et 440-, rejet de {Château, Cadeau, Gratuit} et {Ambition, Vitesse, Dépense}; 453=, indifférence à l'OR.

Au-delà du troisième axe, on trouvera encore matière à interprétation: les modalités principales se disposent de façon cohérente; et les modalités du signalement, surtout les attitudes de vote, suggèrent parfois des conjectures qui embarrasseraient certains des candidats de Mai.



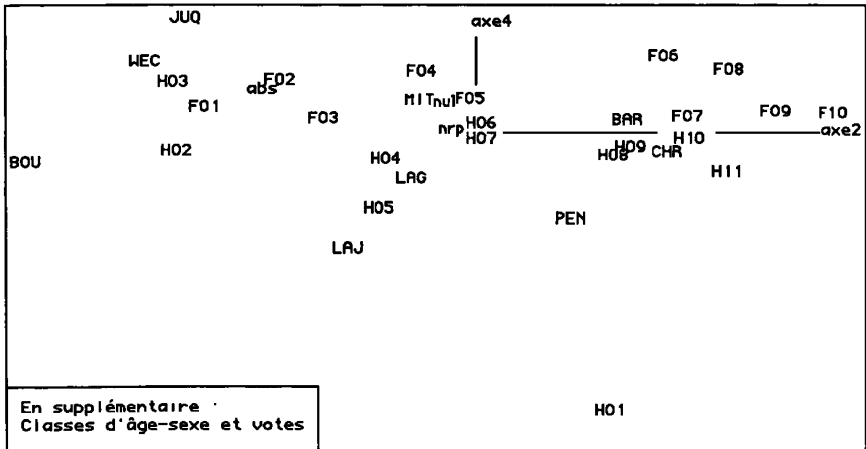
Il faut toutefois prendre garde que les 1140 individus, dont les réponses sont ici traitées, ne sont pas la totalité de ceux interrogés pour les auteurs du questionnaire; et que, même avec un nombre deux fois plus élevé, le terme d'échantillon représentatif ne doit pas faire illusion. Des attitudes complexes face aux idées et à la vie, même exprimées dans des œuvres littéraires, peuvent ne subsister qu'au sein de petits groupes, qui échappent au recrutement des sujets. Dans le cas présent on doit chercher, sur le listage d'analyse des correspondances, si un axe qu'on interprète n'est pas créé par les contributions de quelques individus isolés.

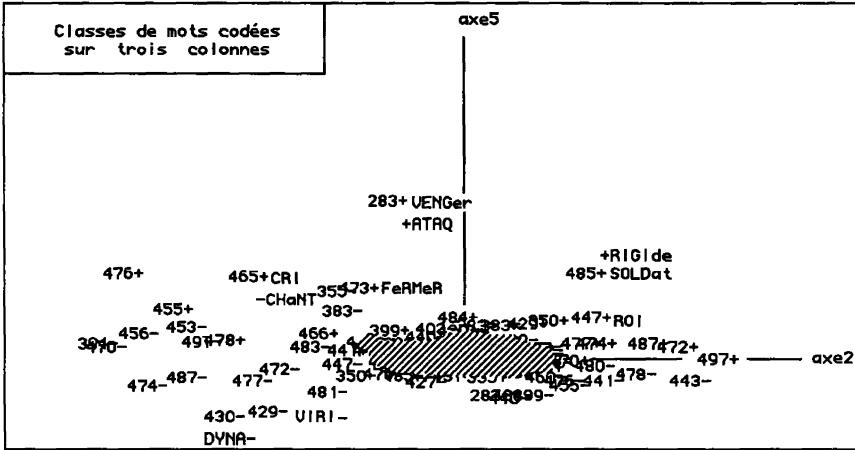




Dans le cas de l'axe 4, ( $\tau_4 = 4.6\%$ ), un quart de l'inertie vient de deux individus; mais la cohérence de la disposition des nombreuses modalités qui s'écartent du côté ( $F_4 < 0$ ) nous donne confiance dans l'interprétation du facteur. Manifestement, il ne s'agit pas de sujets qui s'ingénient à voir la vie en rose, ni d'esthètes.

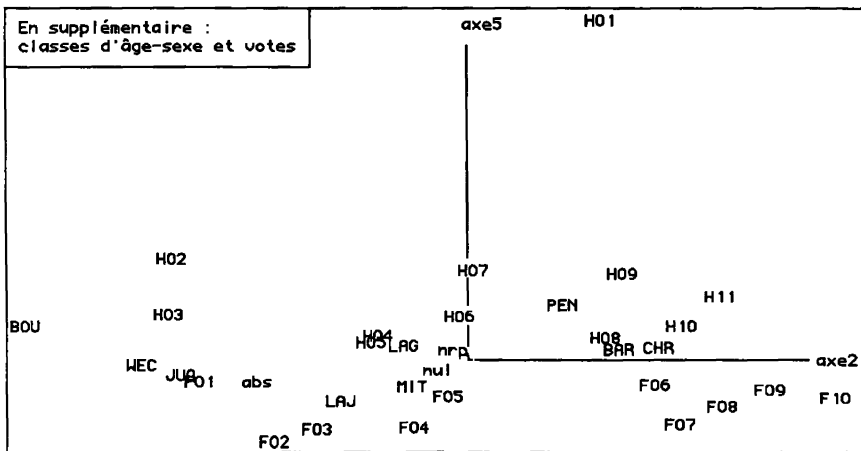
La place d'A. Lajoinie, suivi de J.-M. Le Pen, incite à porter nos regards vers des communes proches de Paris, où celui-ci a moissonné des voix, acquises d'ordinaire au parti de celui-là. À l'opposé, du côté ( $F_4 > 0$ ), on trouve, P. Juquin suivi d'A. Waechter.



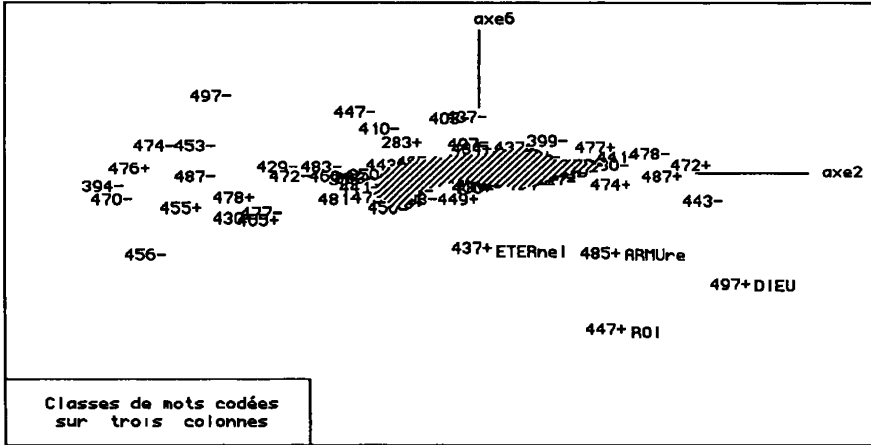


Sur le demi-axe ( $F5 > 0$ ), ( $\tau_5 = 3.5\%$ ), on trouve une conjonction de traits d'agressivité, se projetant aussi bien sur la droite que sur la gauche de l'axe 2. De tous les candidats, Le Pen est le plus engagé de ce côté; à l'opposé, on a notamment le rejet de DYNAMique, de VIRIL.

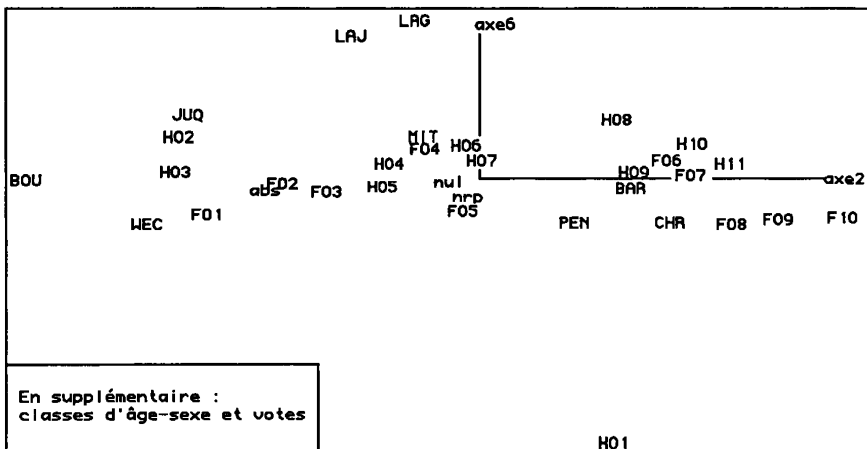
Il est peut-être légitime d'évoquer ici l'idéologie de la *Nouvelle Droite*, qui n'est pas présentement celle d'une masse d'adeptes, mais qui exprime sans doute en quelque manière l'attitude d'un certain nombre d'hommes qui n'ont pas lu Nietzsche.







Avec le demi-axe ( $F_6 < 0$ ), ( $\tau_6 = 3.25\%$ ), c'est assurément l'*Action Française* qu'on croit reconnaître; ou, en général, l'image idéale de la *Monarchie Catholique* et de la *Chevalerie*, (celle-ci, représentée par son armure...). Nos trois jeunes gens de H01 sont présents; mais ils ne sont certainement pas seuls, car nous n'avons pas vu, sur le listage, qu'aucun des 1140 individus apportât de contribution majeure à cet axe, dont l'interprétation doit être considérée avec d'autant plus de confiance que, du côté ( $F_6 > 0$ ), les votes les plus excentrés sont LAJ et LAG.



Il faut seulement prendre garde à la décroissance des valeurs propres: et c'est pourquoi de l'axe 7, ( $\tau_7 = 2\%$ ), nous nous bornerons à dire qu'on y trouve associés, du côté (F7>0), 455+ (ORAG+), 465+ (FOUL+) et 398- (DOUC-).

On terminera ce parcours rapide en attribuant, à chacun des six premiers axes, un ou deux titres qui en rappellent l'interprétation:

- |                                   |                                  |
|-----------------------------------|----------------------------------|
| (F1<0) : certainement ;           | (F1>0) : peut-être ;             |
| (F2<0) : contestation ;           | (F2>0) : religion et morale ;    |
| (F3<0) : centre ;                 | (F3>0) : conjonction ascétique ; |
| (F4<0) : les damnés de la terre ; | (F4>0) : le jardin d'Épicure ;   |
| (F5<0) : agressivité ;            | (F5>0) : refus de la lutte ;     |
| (F6<0) : idéal monarchique.       |                                  |

## 6 Conclusions et perspectives

D'après les facteurs issus de l'analyse objet du §5.4, nous avons effectué une classification ascendante hiérarchique de l'ensemble des 1140 sujets interrogés. Comme de règle, au premier passage, on demande une partition de l'ensemble des sujets en  $n$  classes construite en conservant les  $n-1$  nœuds les plus hauts de la hiérarchie, le nombre  $n$  étant fixé d'après la décroissance des niveaux. Mais ensuite, le programme CAH2 permet d'obtenir rapidement, (i.e. sans refaire de CAH, mais en exploitant les fichiers déjà créés), une partition d'après un ensemble de nœuds spécifiés; nœuds choisis en tenant compte des aides à l'interprétation, et notamment du listage Facor; (dans notre cas, avec 132 variables, le listage Vacor n'est pas directement exploitable; mais on peut étiqueter l'arbre de la CAH d'après des classes de variables). Le programme 'planF' permet de situer rapidement les centres des classes sur tout plan défini par deux axes factoriels.

D'autre part, les résultats obtenus nous donnant confiance dans la validité du codage des notes utilisé jusqu'ici, on peut légitimement tenter d'aller plus loin. Si, en effet, on admet que nos notes recadrées, (cf. §5.1), ont une signification univoque, (l'effet de l'équation personnelle ayant été bien compris), on peut reprendre sur les colonnes ce qui a été fait sur les lignes, soit:

1° Calculer la moyenne  $my_j$ , le maximum  $Mx_j$  et le minimum  $mn_j$  des notes, (notes recadrées entre -1 et +1), figurant dans la colonne  $j$ ;

2° Recadrer la colonne  $j$  suivant la formule:

si  $(k(i, j) > my_j)$  alors  $k(i, j) := (k(i, j) - my_j) / (Mx_j - my_j)$   
 sinon  $k(i, j) := (k(i, j) - my_j) / (my_j - mn_j)$  ;

3° Cumuler les notes ainsi deux fois recadrées suivant la formule même utilisée au §5.3.

L'intérêt de ce double recadrage est de produire des notes  $q^+$  et  $q^-$  équilibrées, qu'il s'agisse d'une classe  $q$  comprenant des mots généralement bien notés ou mal notés; (alors qu'au §5.3 la modalité  $q^-$  est d'autant plus légère que les mots sont plus mal notés par la majorité des sujets). En effet, ici, le point moyen utilisé pour le deuxième recadrage, serait adapté chaque fois au mot considéré. Il faut seulement répéter que, plus on effectue de transformations, plus il convient d'être prudent; c'est-à-dire de vérifier à chaque étape la cohérence des résultats obtenus.

Au point où nous sommes parvenu, il nous paraît démontré que ce vaste corpus de notes, qu'au départ nous ne regardions pas sans suspicion, recèle une représentation riche et nuancée de la diversité des attitudes des sujets interrogés.