

# CAHIER DE ReCHERCHE

SEPTEMBRE 91



N° 18

## ETUDE DE L'OPINION ET ENQUETES DE REFERENCE

Aspects théoriques, méthodologiques et informatiques

(Soutenance : Avril 1988)



Anastassios Iliakopoulos

**CRÉDOC**

**CREDOC**

ETUDE DE L'OPINION  
ET ENQUETES DE REFERENCE

Aspects théoriques, méthodologiques et informatiques

(Soutenance : Avril 1988)

Anastassios Iliakopoulos

Septembre 1991

*Cette thèse de Doctorat en Méthodes Scientifiques de Gestion a été soutenue à l'Université de Paris-Dauphine le 21 Avril 1988 par Anastassios ILIAKOPOULOS, antérieurement à son intégration au CREDOC.*

Université PARIS-DAUPHINE

U.E.R. Sciences des Organisations

Thèse pour l'obtention du titre de Docteur

Méthodes Scientifiques de Gestion

Sujet : Étude de l'Opinion et enquêtes de référence.  
Aspects théoriques, méthodologiques et informatiques.

Directeur de recherche : Jean-Pierre PAGÈS  
Candidat : Anastassios ILIAKOPOULOS

Jury

Président : Bernard ROY  
Suffragants : Jacques ANTOINE  
Pierre CAZES  
Jean-Pierre PAGÈS  
Gilbert SAPORTA  
Michel TENENHAUS

Année de la soutenance : 1988

*"L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs".*

*Je remercie Monsieur le Professeur B. ROY pour l'honneur qu'il me fait en assurant la présidence de ce Jury.*

*Je suis reconnaissant à Messieurs les Professeurs P. CAZES et G. SAPORTA d'avoir accepté la charge de juger ce travail en qualité de rapporteurs.*

*Ma plus profonde reconnaissance va à Monsieur J.P. PAGÈS. Il m'a accueilli au sein du Service d'Études Générales de Protection qu'il dirige au Centre d'Études Nucléaires de Fontenay-Aux-Roses. Il m'a constamment fait bénéficier de son expérience et il a su me communiquer son enthousiasme en m'orientant vers une recherche beaucoup plus ouverte. De plus, il m'a inlassablement incité, encouragé, aidé ; ses remarques constructives se retrouvent dans chaque page de cette thèse.*

*Je remercie Messieurs les Professeurs J. ANTOINE et M. TENENHAUS qui ont bien voulu accepter de participer à ce Jury. C'est à ce dernier que je dois une première découverte de l'Analyse des Données.*

*Je voudrais aussi associer dans un même remerciement amical M.H. BARNY, S. BONNEFOUS, R. MEZIANI, D. VANDERPOOTEN ainsi que tous les autres membres de deux laboratoires : ceux du LSEES à Fontenay-Aux-Roses et ceux du LAMSADE à Paris-Dauphine, qui, au travers des études que nous avons effectuées ensemble ou par l'intérêt et l'amitié qu'ils m'ont témoignés, ont créé des conditions favorables à la réalisation de ce travail.*

## TABLE DE MATIERES

INTRODUCTION . . . . .	1
CHAPITRE I : DE LA CONTROVERSE NUCLEAIRE AU PHENOMENE DE L'OPINION	7
1.1. Les faits : quelques données d'opinion sur le nucléaire .	10
1.2. Approche mécaniste : l'observateur et la relation sujet/objet . . . . .	15
1.3. Approche constructionniste : la prise en compte du processus de communication . . . . .	17
1.4. Vers une économie des rôles . . . . .	19
Bibliographie . . . . .	22
CHAPITRE II : APPROCHE SYSTEMIQUE DE L'ANALYSE FACTORIELLE ET ENQUETES DE REFERENCE : DU MODELE STIMULUS-REPONSE AUX ANALYSES FACTORIELLES . . . . .	23
2.1. En suivant Charles Spearman : approche déterministe des enquêtes de référence . . . . .	24
2.2. En suivant Karl Pearson : approche statistique des enquêtes de référence . . . . .	28
2.3. Réalisation d'une enquête de référence . . . . .	33
Bibliographie . . . . .	37
CHAPITRE III : ENQUETES DE REFERENCE ET COMMUNICATION ENTRE ENQUETES . . . . .	38
3.1. Présentation critique de trois méthodes . . . . .	40
3.1.1. L'approche de Jöreskog . . . . .	41
3.1.2. Analyse Conjointe de Plusieurs Tableaux de Données	47
3.1.3. L'analyse de redondance . . . . .	50
3.2. Méthodologie proposée . . . . .	52
3.2.1. Notion de passerelle . . . . .	52
3.2.2. Sélection d'un sous-ensemble optimal de variables	54
3.2.3. Elimination de variables . . . . .	60
3.2.4. Technique par régression (méthode de la "greffe")	67
3.2.5. Technique par rotations procrustéennes . . . . .	75
Bibliographie . . . . .	88

<b>CHAPITRE IV : SUR LE TRAITEMENT INFORMATIQUE DES ENQUETES D'OPINION</b>	91
4.1. Création du fichier de travail . . . . .	97
4.2. Tris à plat . . . . .	100
4.3. Tris croisés . . . . .	100
4.4. Analyse en Composantes Principales . . . . .	106
4.4.1. Présentations de la matrice des corrélations- Choix des variables actives . . . . .	108
4.4.2. Choix du nombre d'axes à retenir . . . . .	111
4.5. Analyse des Correspondances . . . . .	116
4.6. Classification Automatique . . . . .	117
4.7. Rotations Analytiques . . . . .	119
4.8. Classification de variables . . . . .	123
4.9. Stabilité des points moyens . . . . .	124
Bibliographie . . . . .	129
<b>CHAPITRE V : EVALUATION DE LA DEMARCHE A L'AIDE D'EXEMPLES</b> . . . .	131
5.1. Le repère de référence (enquête de Mars 1981) . . . . .	132
5.2. Un éclairage sur l'action militante et la protection sociale . . . . .	141
5.3. Sur l'évolution des structures de l'opinion . . . . .	158
5.4. Structures de l'opinion et perception des risques . . . . .	172
<b>EPILOGUE</b> . . . . .	180

## INTRODUCTION

Quand on se préoccupe du phénomène de l'opinion pris dans toute sa généralité, les sociétés démocratiques modernes peuvent être assimilées à des systèmes dynamiques régulés par les conflits. Conséquence de la séparation et du pluralisme des pouvoirs, ces conflits ne sont pas seulement une affaire d'acteurs ; ils impliquent le public dès que, prenant de l'ampleur, ils apparaissent sur la scène de la presse et des médias. Ainsi se trouve posé devant l'opinion un problème n'ayant pu être résolu dans un cadre plus restreint.

L'opinion est donc une force qui ne se manifeste pas seulement à intervalles réguliers lors des élections ou des référendum. Dans les conflits, elle exerce son influence à travers les sondages ; ces modes nouveaux de consultation se sont développés de façon considérable ces dernières années, en France en particulier.

Avec les élections, la remise en cause affecte avant tout le système des acteurs politiques qui sont l'une des pierres angulaires dans nos états modernes. Mais, avec les sondages, cette remise en cause peut concerner directement tout acteur de la vie économique et sociale qui se trouve, brutalement, confronté au public sous les projecteurs de l'actualité. Comme pour l'acteur politique, son avenir peut alors se jouer. Les responsables des grandes entreprises et organisations doivent donc intégrer dans leur stratégie une nouvelle composante de l'opinion, composante qui est loin d'être celle qui était prise en compte, récemment encore, dans les relations publiques et dans les études de marketing. On ne se situe pas ici dans le cadre de la rationalité classique ; aucune relation mécanique n'existe, a priori, entre, d'un côté, les qualités objectives des individus et des objets sur lesquels ils se prononcent, et, de l'autre, l'opinion qu'ils expriment. On se trouve dans le cadre d'une "rationalité de conflit" où les significations se construisent au fur et à mesure, au gré des passes d'armes ou des alliances.

Le phénomène de l'opinion dont on traite, fait référence à un processus dynamique d'apprentissage. Le recours au langage permet aux forces de se manifester et ainsi, en autorisant les ajustements, d'assurer la convergence vers un état d'équilibre. Précisons. Les avis et les arguments produits face au public par les acteurs rendent patente la négociation qui se déroule en permanence entre les organisations antagonistes ; à travers ses porte-parole, chacune d'entre elles cherche à améliorer ou à défendre sa position. Les individus procèdent de façon équivalente : donner son opinion sur un sujet de grande actualité, c'est aussi faire un choix stratégique dans la relation que l'on développe avec autrui. Il s'agit là encore de conforter une position ; mais le bénéfice tiré n'est pas évident. Moyen pour exprimer sa différence, l'opinion est certainement un élément qui joue un rôle important dans la construction des personnalités.

*Comment alors traduire sous forme d'outil d'analyse cette façon d'appréhender le phénomène de l'opinion ? De façon plus précise : comment procéder à l'aide d'une enquête pour obtenir les caractéristiques de ce système, en perpétuelle évolution, où communiquent avec des logiques différentes les acteurs porte-parole des institutions et le public ? Comment, dans un sondage, utiliser ces caractéristiques pour analyser plus finement les attitudes des gens dans des circonstances particulières ? Enfin, comment exploiter toutes ces informations pour envisager des stratégies de communication qui permettent d'améliorer sa position dans les rapports de force ?*

Le système d'enquêtes sur "les structures de l'opinion publique" dont la promotion est assurée, à l'heure actuelle, par l'association Agoramétrie (association loi de 1901, créée à l'initiative du Commissariat à l'Energie Atomique et d'Electricité de France), a été mis en place pour tenter de répondre aux problèmes soulevés dans les questions précédentes. Il s'agissait alors pour nous, partant de l'existant, d'approfondir les fondements théoriques de ces enquêtes, de développer une méthodologie permettant d'y avoir recours effectivement comme cadre de référence et de mettre au point les outils informatiques sans lesquels rien ne serait aujourd'hui opérationnel.

Pour atteindre ces objectifs, on s'est efforcé d'allier modernisme et simplicité au niveau de l'approche et du langage, et pragmatisme au niveau des méthodes et des outils. En adoptant un langage qui peut être considéré comme "systémique", on a d'abord, en échappant aux dialectes du psychologue et du sociologue, donné un caractère général aux schémas de base élaborés ; puis, on s'est réservé la possibilité de circuler entre une première approche déterministe du phénomène de l'opinion (modélisation) et une seconde approche entièrement statistique (mise en évidence des structures). C'est, alors, dans l'esprit et le respect des principes de l'Analyse des Données, qu'ont pu être abordés, sans fioritures, les aspects méthodologiques ; on a ainsi évité de tomber dans les finesses excessives caractéristiques des pratiques anglo-saxonnes. Enfin, les outils informatiques ont été conçus avec le souci premier de mettre en valeur le système d'enquêtes en s'adaptant aux réalités informatiques actuelles ; on a préféré, ici, gagner en rapidité et en esthétique au détriment de la généralité.

Dans le chapitre I on introduit les premiers éléments d'une théorie qui est celle à laquelle on se réfère quand on parle du phénomène de l'opinion. Pour aboutir à un schéma qui met en place les concepts de base (acteurs, public, environnement) et les relations qu'ils entretiennent les uns avec les autres, on part d'une analyse des opinions sur le programme nucléaire : quels sont les faits importants en 1987 ? que s'est-il passé depuis dix ans ?

L'approche mécaniste, consistant à traiter séparément les objets de discussion et les individus ou groupes sociaux qui en discutent, et l'approche constructionniste, où individus et objets, en dualité, ne peuvent être analysés que simultanément, sont successivement mises à l'épreuve. C'est ainsi que, de proche en proche, on aboutit à un schéma similaire à l'un de ceux auxquels fait référence l'économie du marché : "le schéma de l'économie des rôles". Avec ce schéma, l'élément (l'individu, l'acteur) et le tout (la société dans son ensemble, l'environnement) entrent en résonance à travers toute une série de chocs de natures différentes (conflits entre éléments, conflits entre acteurs et public) qui se propagent de bas en haut et de haut en bas en traversant les frontières correspondant aux différents niveaux d'organisation (sous-systèmes "des acteurs" et "du public"). L'opinion est donc traitée comme

une force. Comme en économie, la difficulté est ici de raccrocher le *micro* (le comportement de chacun des individus et de chacun des acteurs) au *macro* (l'opinion "publique").

Ayant posé un schéma, il reste à construire le modèle. Dans le chapitre II on montre comment on peut adapter le vieux modèle "Stimulus (Personnalité) Réponse" (modèle  $S \rightarrow R$  ou  $S \xrightarrow{P} R$ ) dans le cadre d'une approche constructionniste. Ce n'est plus l'individu, soumis au stimulus, qui est alors considéré comme système, mais c'est ce nouvel objet : l'interaction, qu'il forme avec le stimulus.

Après linéarisation du modèle qui prend en charge l'interaction, on est conduit à retenir l'Analyse en Composantes Principales comme technique factorielle d'extraction des paramètres de ce modèle simplifié, qui n'est autre que le modèle factoriel des psychomètres. On respecte ainsi le principe de dualité introduit précédemment. Comme l'homme crée le symbole (signification du stimulus) qui l'aliène en retour (structuration de sa pensée), il est logique, en effet, de décrire avec les mêmes paramètres (les dimensions principales) l'univers des stimuli et la population des individus.

La discussion sur le modèle factoriel et les techniques d'Analyse Factorielle qui permettent de le mettre en oeuvre conduit naturellement à introduire la notion d'enquête de référence qui est au centre de la recherche entreprise.

Les enquêtes de référence dont on traite fournissent des indicateurs globaux d'opinion (les dimensions de la différenciation extraites par l'Analyse en Composantes Principales) qui permettent de se resituer dans un cadre général quand on aborde l'étude d'un problème spécifique (par exemple, la perception des risques). Il faut donc faire passer l'information que représentent ces indicateurs d'une enquête à l'autre. C'est à ce problème de communication entre enquêtes qu'est consacré le chapitre III où se concrétisent, à travers la méthodologie qui y est développée, toutes les idées initiales sur l'opinion.

Pour mettre en évidence l'efficacité et l'originalité des propositions qui sont faites en matière de communication entre enquêtes, un

bilan critique des méthodes les plus connues est d'abord effectué. On insiste, en particulier, sur les travaux de K. Jöreskog, références classiques dans le monde anglo-saxon, qui donnent une bonne idée de la façon d'aborder un problème en statistique inférentielle quand on se place dans une optique, non pas "exploratoire", mais "confirmatoire". C'est dans cette dernière optique que l'on va se placer, mais on n'hésitera pas à se débarrasser de toute référence au modèle probabiliste pour rester dans l'esprit de l'Analyse des Données.

La méthodologie développée a pour objectif la reconstruction, à l'intérieur d'une enquête particulière, des indicateurs factoriels de l'enquête de référence. Deux techniques sont proposées : l'une s'inspire de la régression, l'autre fait appel aux rotations procrustéennes. Elles ne peuvent être utilisées à bon escient que si on introduit des hypothèses sur le phénomène étudié : les idéologies sont-elles structurées de manière équivalente dans des populations différentes saisies au même moment (synchronie) ? le sont-elles, dans une même population, quand celle-ci est suivie au cours du temps (diachronie) ? Comment tenir compte du fait que certains objets de discussion changent de signification d'un groupe social ou d'une époque à l'autre ? La qualité du recalage sur les structures de référence, quelle que soit la technique utilisée, va reposer sur tout un travail de préparation. Il s'agit d'extraire de l'échantillon des variables un sous-ensemble, qu'on appelle passerelle, qui puisse être considéré comme représentatif de l'univers des variables et qui soit cohérent avec les hypothèses faites sur le phénomène.

Les méthodes introduites n'ont intérêt que dans la mesure où elles conduisent sur le plan informatique à des logiciels performants. Ces logiciels doivent s'intégrer dans un ensemble de programmes adapté au traitement de grosses enquêtes. Dans le chapitre IV on décrit, dans les grandes lignes, le logiciel de traitement d'enquêtes sur micro-ordinateur mis au point. On a cherché à construire un outil très performant et produisant, avec un double souci d'esthétique et d'économie, des documents statistiques de qualité. En même temps, on a fait en sorte que l'utilisateur soit incité à se plier à une certaine discipline qui reflète notre conception du traitement d'enquêtes. Cette conception se traduit, au niveau des analyses statistiques, par diverses procédures originales qui font l'objet de développements particuliers.

Il reste à prouver l'efficacité de la démarche. Le modèle initial est-il réaliste ? Que valent les méthodes qui ont été développées ? Dans le chapitre V, qui peut être considéré comme un bilan critique de la recherche, on démontre, sur trois exemples, la pertinence de l'idée initiale (l'opinion est un fait social qui se construit dans un débat) et du schéma, du modèle et des méthodes mis au point. Le pouvoir explicatif des indicateurs globaux que sont les dimensions factorielles de l'enquête de référence construite à partir des conflits s'avère dans les trois exemples remarquable.

## CHAPITRE I

### DE LA CONTROVERSE NUCLEAIRE AU PHENOMENE DE L'OPINION

#### INTRODUCTION

Quand on cherche à bâtir un système de suivi des opinions, il faut d'abord préciser ce qu'est l'opinion dont il s'agit. Pour le dictionnaire alphabétique et analogique de la langue française, Le Robert, l'opinion est la "manière de penser, de juger sur tel ou tel sujet, qu'il s'agisse d'une attitude de l'esprit qui tient pour vraie une assertion, ou d'une assertion, d'un système d'assertions que l'esprit accepte ou rejette (généralement en acceptant une possibilité d'erreur)". Cette définition correspond bien à la conception qu'a de l'opinion le chercheur en psychologie sociale dont les travaux restent, en général, inconnus du public ; celui-ci s'efforce de cerner, en construisant les instruments adéquats, les manières de juger et de penser de l'individu et cela compte tenu de son statut social. Elle est incomplète pour le politologue qui s'intéresse avant tout aux acteurs politiques et aux grands débats nationaux et internationaux dans lesquels ceux-ci interviennent ; participant à la vie médiatique, il considère implicitement l'opinion comme une expression du principe du peuple souverain, exerçant en permanence son influence grâce à ce mode moderne de consultation populaire qu'est le sondage. Cette différence de définition transparait dans les questionnaires des enquêtes qui sont effectuées.

Dans les enquêtes dont les résultats sont publiés dans la presse, la façon de poser la question est généralement déterminante : un pourcentage dépend fortement du libellé retenu<sup>1</sup>. Il s'agit ici, non pas d'établir une vérité, mais simplement d'indiquer une tendance qui, si elle fait réfléchir le décideur, peut surtout lui fournir un argument (une légitimité) pour envisager un ensemble de mesures. Un pourcentage isolé qui n'est pas mis en perspective (comparaison avec d'autres pourcentages qui peuvent correspondre à d'autres questions, d'autres groupes sociaux ou d'autres époques) ne peut servir à la compréhension du phénomène ; n'étant pas un instrument de connaissance<sup>2</sup>, il ne peut donc être qu'un instrument de pouvoir. On voit donc bien ici apparaître l'opinion comme une force qui peut se révéler déterminante pour l'action ; elle est cette opinion qui "n'existe pas" dont parle P. Bourdieu [Bourdieu 1973].

Dans le baromètre Agoramétrie, c'est bien de cette "opinion-force" dont on traite apparemment puisque toutes les questions sont posées sous forme de propositions-référendum : il faut rétablir la peine de mort ; la réduction des dépenses militaires s'impose... Mais on n'est pas ici à la recherche d'une légitimité pour agir. On cherche à comprendre le phénomène social auquel renvoie l'opinion-force : que veut-on dire quand on s'exprime sur tel sujet ? que dit-on quand on s'exprime sur tous les sujets à la fois ? Aussi ne procède-t-on pas en "ordre dispersé" mais de façon globale. La question n'est pas isolée ; elle est élément d'un corpus qui se veut représentatif de l'ensemble des problèmes auxquels se trouve confrontée à un moment donné une société. Et c'est ainsi, en considérant la question comme élément d'un tout, que l'on échappe au problème du libellé. Les structures dégagées à partir d'un questionnaire particulier ne seront pas sensibles à des fluctuations dans la formulation des questions ; et même plus, caractéristiques du tout, elles seront quasi-indépendantes des questions posées.

---

<sup>1</sup> Dans les enquêtes réalisées par Agoramétrie en 1977 et 1978, par exemple, deux libellés différents ont été utilisés pour traduire l'adhésion des français à l'énergie solaire : en 1977, 46% ont été d'accord avec la proposition "il faut dépenser beaucoup d'argent pour l'énergie solaire" alors qu'en 1978, avec la proposition "il faut développer au maximum l'utilisation de l'énergie solaire", on en a trouvé 81%.

<sup>2</sup> sauf dans les rares cas où on cherche à prévoir les résultats d'un suffrage (élections, référendum).

Avec l'opinion force des politologues telle qu'on l'aborde à Agoramétrie, on est donc proche des sujets dont traite habituellement le chercheur en psychologie sociale. Mais on ne s'intéresse pas à la mesure d'une attitude particulière sur un thème quelconque<sup>1</sup> ; on étudie les déterminants des attitudes sur les thèmes d'opinion conflictuels, en général, qui sont toujours d'ordres psychologique et social à la fois.

Ayant précisé à quelle opinion on s'intéresse, il reste pour concevoir le système de suivi des opinions à en dire davantage sur les mécanismes mis en jeu. A quel phénomène social l'opinion-force fait référence ?

Dans ce chapitre on montre comment, en s'appuyant sur des données d'opinion sur le programme nucléaire, on est conduit à assimiler ce phénomène à un "système" représentable sous la forme d'un schéma simple. Pour aboutir à ce "schéma de l'économie des rôles", on a procédé par étape à partir d'une première esquisse de modèle, en faisant disparaître au fur et à mesure les insuffisances mises en évidence en se confrontant aux faits. On pourra ainsi expliciter les objectifs des enquêtes et ramener l'analyse de l'opinion à celle d'un modèle stimulus-réponse tel qu'il est développé dans le deuxième chapitre.

---

<sup>1</sup> Citons par exemple l'échelle de distance sociale mise au point par Bogardus [Bogardus 1933]. Si X désigne une race, une ethnique, un peuple (Arménien, Américain (blanc), Chinois...), on pose successivement les questions suivantes qui s'ordonnent suivant une "échelle de Guttman" (réponses par "oui" ou "non") :

- Je me marierais bien avec X
- J'accepterais comme amis intimes X
- J'accepterais comme voisins X
- Je travaillerais bien dans le même bureau que X
- Je n'accepterais que d'avoir des relations verbales avec X
- J'accepterais seulement comme visiteurs de mon pays X
- J'exclurais de mon pays X

Ces questions permettent d'associer à chaque interviewé une note qui traduit la sympathie manifestée (l'attitude) pour X.

### 1.1. Les faits : quelques données d'opinion sur le nucléaire

Les données d'enquête dont on dispose<sup>1</sup> permettent d'examiner l'opinion sur le programme nucléaire civil dans deux perspectives :

- celle de la synchronie : quelle est la cohérence entre les positions prises sur des sujets divers ? Comment se différencient dans leurs opinions les différents groupes sociaux ?

- celle de la diachronie : comment évolue l'opinion en général ou dans certains groupes sociaux ? Comment évoluent les "structures" qui seraient à un moment donné caractéristiques de l'opinion ?

#### Synchronie

Dans le tableau ci-après, issu de l'enquête Agoramétrie réalisée en Juin 1987, apparaissent les plus fortes corrélations qu'entretient le thème "il faut continuer à construire des centrales nucléaires" avec les autres thèmes de conflit considérés dans la même enquête (à tous ces thèmes les réponses se font sur une échelle "d'accord" en cinq paliers).

Pour la force de frappe	0.284
Exporter les armes françaises	0.255
La privatisation, c'est bien	0.252
Réduire les dépenses militaires	-0.246
Soutenir les écologistes	-0.242
Soutenir les pacifistes	-0.215
Industrie chimique polluante	-0.213
Pour des télévisions privées	0.199
Nécessité d'expérimenter sur les animaux	0.197
Vérité cachée en France sur Tchernobyl	-0.192
Publicité indispensable	0.191

L'opinion sur les centrales nucléaires est donc liée à celle sur la force de frappe, les privatisations, les écologistes, les télévisions privées et la publicité...

Dans le tableau de la figure 1.1 on voit apparaître les liaisons, en 1987, entre l'opinion sur les centrales nucléaires et les facteurs région, habitat, sexe, âge, nombre d'enfants à charge, religion,

---

<sup>1</sup> L'opinion sur le nucléaire civil est suivie depuis environ les années 60 [Fourgous et al 1980] et de façon régulière par EDF et Agoramétrie depuis les années 75.

Thème n° 4 : Construire des centrales nucléaires													
1 PAS DU TOUT D'ACCORD	509	34.6	61.4		■■■■■■■■■■								
2 PAS TELLEMENT D'ACCORD	394	26.8	19.9		■■■■■■■■								
3 PEUT-ETRE D'ACCORD	292	19.9	11.9		■■■■■■								
4 BIEN D'ACCORD	175	11.9	6.8		■■■■								
5 ENTIEREMENT D'ACCORD	100	6.8			■■								
REGION	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
REGION PARISIENNE	81	78	60	28	19	266	305	293	226	105	71	1000	2.35
NORD	35	29	21	14	10	109	321	266	193	128	92	1000	2.40
EST	48	24	31	22	7	132	364	182	235	167	53	1000	2.36
BASSIN PARISIEN	89	91	54	36	23	293	304	311	184	123	78	1000	2.36
OUEST	75	45	45	21	13	199	377	226	226	106	65	1000	2.26
CENTRE-EST	63	40	23	20	7	153	412	261	150	131	46	1000	2.14
SUD-OUEST	63	40	24	16	10	153	412	261	157	105	65	1000	2.15
MEDITERRANEE	55	47	34	18	11	165	333	285	206	109	67	1000	2.29
TOTAL	509	394	292	175	100	1470							
chi-2 = 28.46 ddl = 28 Proba = 0.44043													
HABI HABITAT (4 MODALITES	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
- DE 2000 H.	121	92	73	54	28	368	329	250	198	147	76	1000	2.39
- DE 100000 H.	147	113	92	55	19	426	345	265	216	129	45	1000	2.26
+ DE 100000 H.	164	130	84	41	36	455	360	286	185	90	79	1000	2.24
PARIS	77	59	43	25	17	221	348	267	195	113	77	1000	2.30
TOTAL	509	394	292	175	100	1470							
chi-2 = 13.56 ddl = 12 Proba = 0.32953													
SEXE	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
HOMME	204	164	151	113	68	700	291	234	216	161	97	1000	2.54
FEMME	305	230	141	62	32	770	396	299	183	81	42	1000	2.07
TOTAL	509	394	292	175	100	1470							
chi-2 = 56.06 ddl = 4 Proba = 2.0E-011													
AGE (5 MODALITES)	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
18 A 24 ANS	73	62	52	28	10	225	324	276	231	124	44	1000	2.29
25 A 34 ANS	128	103	71	37	20	359	357	287	198	103	56	1000	2.21
35 A 49 ANS	118	94	66	41	33	352	335	267	188	116	94	1000	2.37
50 A 64 ANS	108	69	67	37	28	309	350	223	217	120	91	1000	2.38
65 ANS ET PLUS	82	66	36	32	9	225	364	293	160	142	40	1000	2.20
TOTAL	509	394	292	175	100	1470							
chi-2 = 20.60 ddl = 16 Proba = 0.19444													
ENFANTS A CHARGE	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
AUCUN	313	237	160	112	54	876	357	271	183	128	62	1000	2.27
UN ENFANT	83	67	57	21	20	248	335	270	230	85	81	1000	2.31
DEUX ENFANTS	67	63	45	25	14	214	313	294	210	117	65	1000	2.33
TROIS ENFANTS ET PLUS	46	27	30	17	12	132	348	205	227	129	91	1000	2.41
TOTAL	509	394	292	175	100	1470							
chi-2 = 11.97 ddl = 12 Proba = 0.44781													
RELIGION	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
RELIGION	104	63	45	31	15	258	403	244	174	120	58	1000	2.19
CATHOLIQUE NON PRATIQUANT	216	166	132	67	45	626	345	265	211	107	72	1000	2.30
PRATIQUANT OCCASIONNEL	97	87	77	49	23	333	291	261	231	147	69	1000	2.44
PRATIQUANT REGULIER	50	49	24	18	13	154	325	318	156	117	84	1000	2.32
AUTRES CAS	42	29	14	10	4	99	424	293	141	101	40	1000	2.04
TOTAL	509	394	292	175	100	1470							
chi-2 = 21.10 ddl = 16 Proba = 0.17456													

FIGURE 1.1 : Tris croisés d'un thème de conflit (Suite...)

Construire des centrales nucléaires													
DIPLOME (6 MODALITES)	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
AUCUN DIPLOME	123	70	45	25	13	276	446	254	163	91	47	1000	2.04
C.E.P.	71	48	48	25	11	203	350	236	236	123	54	1000	2.30
C.A.P.	101	63	49	31	19	263	384	240	186	118	72	1000	2.25
BREVET	82	79	51	39	27	278	295	284	183	140	97	1000	2.46
BAC	58	56	39	29	14	196	296	286	199	148	71	1000	2.41
SUPERIEUR	74	78	60	26	16	254	291	307	236	102	63	1000	2.34
TOTAL	509	394	292	175	100	1470							
chi-2 =	34.93	ddl = 20		Proba = 0.02051									
CSP INTERVIEWE (11 MOD.)	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
AGRICULTEURS	14	15	8	8	2	47	298	319	170	170	43	1000	2.34
COMMERCANTS ET PATRONS	30	28	17	11	9	95	316	295	179	116	95	1000	2.38
CAD.SUP.,PROF. LIBERALES	15	27	14	9	9	74	203	365	189	122	122	1000	2.59
ENSEIGNANTS ET ETUDIANTS	42	44	39	25	6	156	269	282	250	160	38	1000	2.42
PROFESSIONS INTERMEDIAIRE	34	18	20	5	6	83	410	217	241	60	72	1000	2.17
EMPLOYES	89	69	51	28	17	254	350	272	201	110	67	1000	2.27
OUVRIERS ET CONTREMAITRES	60	38	36	22	9	165	364	230	218	133	55	1000	2.28
OUVRIERS SPECIALISES	26	16	12	2	6	62	419	258	194	32	97	1000	2.13
RETRAITES OUVRIERS	1	2	1	1	0	5	200	400	200	200	0	1000	2.40
AUTRES RETRAITES	104	69	46	39	20	278	374	248	165	140	72	1000	2.29
SANS PROFESSION	94	68	48	25	16	251	375	271	191	100	64	1000	2.21
TOTAL	509	394	292	175	100	1470							
chi-2 =	42.73	ddl = 40		Proba = 0.35475									
POLITIQUE	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
EXTREME GAUCHE	8	3	1	1	1	14	571	214	71	71	71	1000	1.86
GAUCHE	200	127	80	40	25	472	424	269	169	85	53	1000	2.07
CENTRE	75	76	59	31	13	254	295	299	232	122	51	1000	2.33
DROITE	64	87	65	54	44	314	204	277	207	172	140	1000	2.77
EXTREME DROITE	4	4	1	1	4	14	286	286	71	71	286	1000	2.79
INDIFFERENT	83	48	46	24	8	209	397	230	220	115	38	1000	2.17
NON REPONSE	75	49	40	24	5	193	389	254	207	124	26	1000	2.15
TOTAL	509	394	292	175	100	1470							
chi-2 =	98.96	ddl = 24		Proba = 4.5E-011									
PROXIMITE POLITIQUE	1	2	3	4	5	TOTAL	1	2	3	4	5	TOTAL	MOY.
EXTREME GAUCHE	7	2	1	1	0	11	636	182	91	91	0	1000	1.64
P.C.F.	34	6	11	7	4	62	548	97	177	113	65	1000	2.05
P.SOCIALISTE	167	125	73	44	21	430	388	291	170	102	49	1000	2.13
RADICAUX GAUCHE	13	7	5	4	2	31	419	226	161	129	65	1000	2.19
ECOLOGISTES	35	20	6	4	1	66	530	303	91	61	15	1000	1.73
U.D.F.	33	44	41	27	20	165	200	267	248	164	121	1000	2.74
R.P.R.	39	53	38	38	28	196	199	270	194	194	143	1000	2.81
FRONT NATIONAL	13	9	8	3	8	41	317	220	195	73	195	1000	2.61
AUCUN	115	87	77	30	11	320	359	272	241	94	34	1000	2.17
NON REPONSE	53	41	32	17	5	148	358	277	216	115	34	1000	2.19
TOTAL	509	394	292	175	100	1470							
chi-2 =	126.46	ddl = 36		Proba = 5.4E-012									

Enquête Agoramétrie 1987 : n=1470 interviewés.

Croisements du thème de conflit :

"il faut continuer à construire des centrales nucléaires"  
avec les principales variables du signalétique.

FIGURE 1.1 : (...Suite)

diplôme, profession, tendance politique et préférences partisans. Les femmes et la gauche sont plus défavorables au nucléaire que les hommes et la droite, respectivement. Les liaisons ayant évolué au cours du temps, pour hiérarchiser les facteurs, il faut tenir compte des résultats des enquêtes obtenus depuis 10 ans. Les variables socio-démographiques et socio-culturelles les plus explicatives de l'attitude sur le nucléaire sont par ordre d'importance :

- la tendance politique,
- le sexe,
- l'âge : l'écart entre les moins de 35 ans qui sont généralement les plus opposés au nucléaire et les plus de 50 ans était de l'ordre de 15% à 20% ; depuis Tchernobyl, cet écart s'est réduit,
- la profession : les enseignants qui étaient avant 1981 les plus opposés, ont effectué un chassé-croisé avec les retraités qui étaient avant parmi les plus favorables (voir figure 1.2),
- la religion : les athées, surtout avant 1982, étaient plus opposés au nucléaire que les catholiques pratiquants,
- l'habitat : si on ne constate pas de différences nettes entre habitat rural et grandes villes, on note un effet "site nucléaire" important : les personnes habitant autour des centrales sont en général plus favorables au nucléaire.

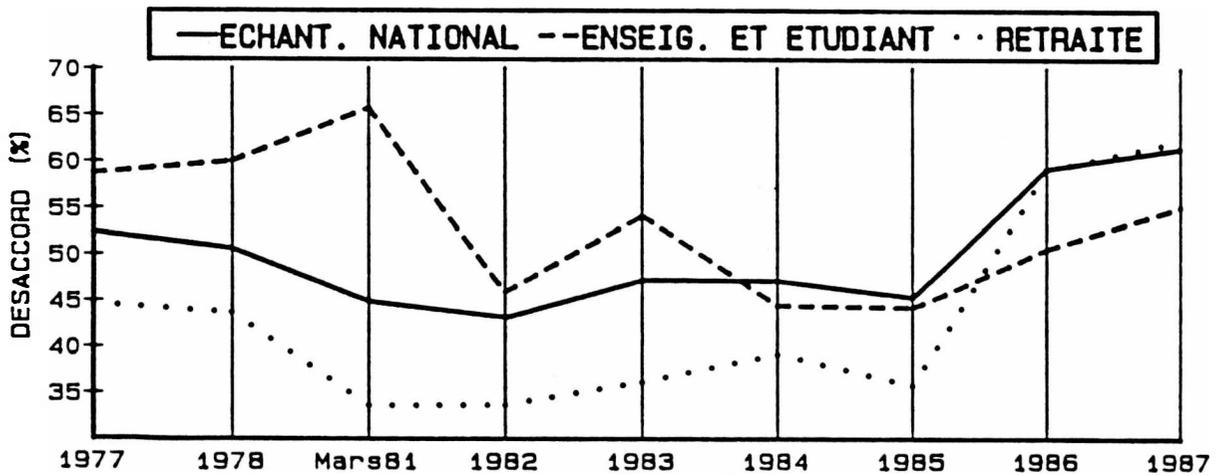
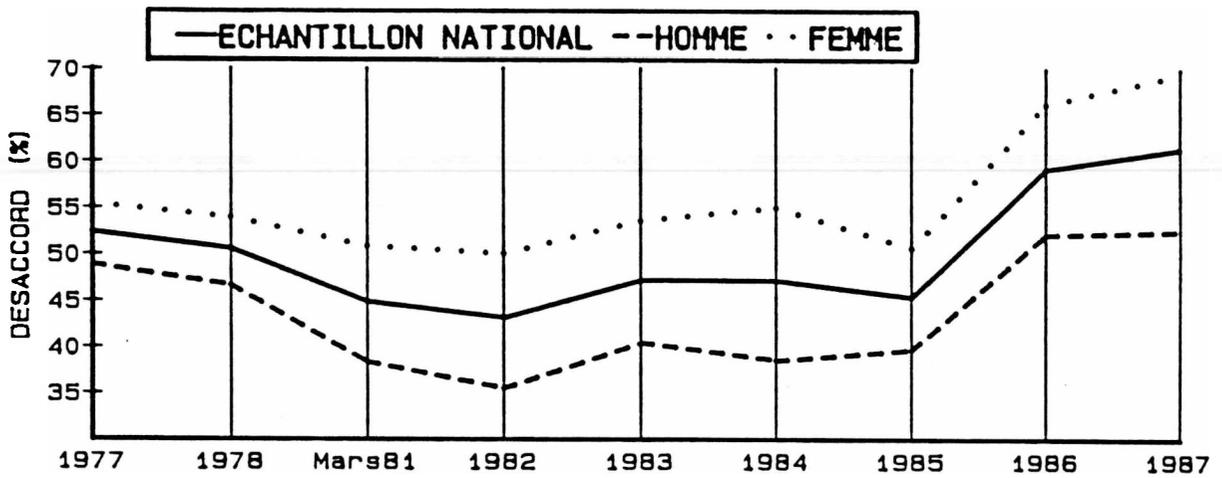
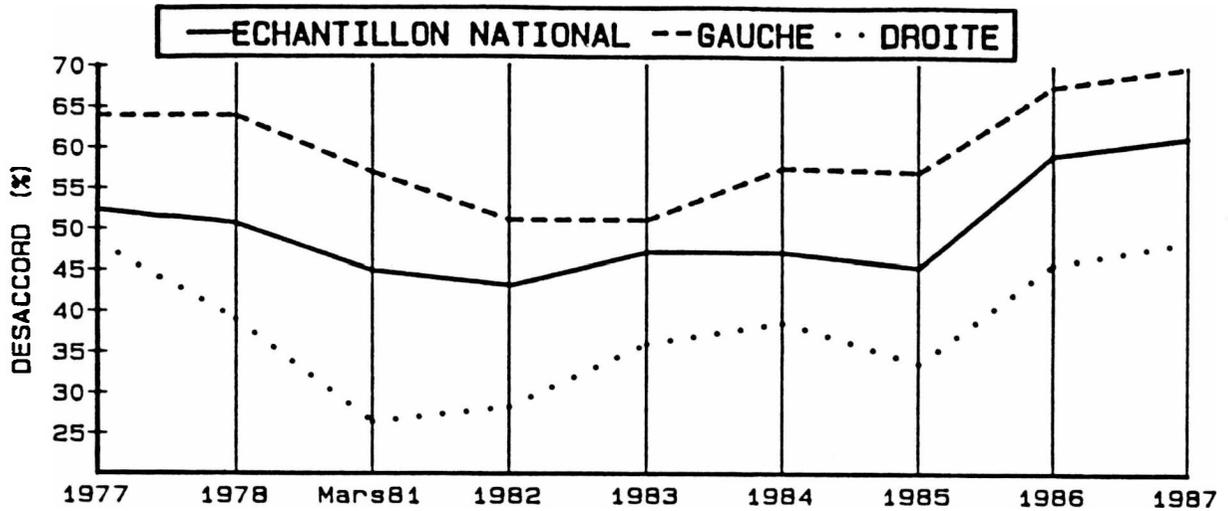
### Diachronie

Si on résume, dans l'évolution des opinions de 1977 à 1981, on peut distinguer trois périodes (voir figure 1.2).

- La première va de 1977 jusqu'aux élections présidentielles de 1981. L'image du programme nucléaire va s'améliorant "en moyenne", les opposants au nucléaire devenant minoritaires<sup>1</sup>. Deux faits majeurs permettent d'expliquer cette évolution : les manifestations de Creys-Malville et surtout la crise de l'énergie.
- La seconde va de mai 1981 à avril 1986. La relative stabilité des pourcentages globaux cache une réorganisation en profondeur des opinions : amélioration très nette de l'image des centrales chez les intellectuels, venant compenser une certaine désaffection des

---

<sup>1</sup> On assimile les résultats du sondage, ici, à ceux d'un référendum.



Neuf enquêtes nationales (1977 à 1987) d'Agoramétrie.

Evolution du pourcentage de désaccord avec la proposition : "il faut continuer à construire des centrales nucléaires"

FIGURE 1.2 : Evolution des pourcentages

groupes, comme les retraités, traditionnellement situés à droite. L'acte symbolique que constitue l'abandon de Plogoff, d'un côté, et les changements institutionnels, de l'autre, ont certainement facilité ces mouvements contradictoires de groupes sociaux de sensibilités opposées.

- La troisième, qui commence en mai 1986, fait suite à l'accident de Tchernobyl. La détérioration de l'image du nucléaire dès juin 1986 (avancée de plus de 10% des avis défavorables à la construction de nouvelles centrales) s'est maintenue depuis ; elle affecte uniquement le programme nucléaire civil : les images des experts et des institutions n'ont pas souffert de l'accident. Le repli ne s'est pas accompagné d'un véritable effet de structure ; il a permis seulement de parfaire le réajustement des positions qu'avait déclenché la victoire de la gauche en 1981.

Il reste à organiser en un tout cohérent les informations réunies, qui peuvent paraître paradoxales : comment tirer un modèle de cette complexité ? Faisant émerger les éléments théoriques d'une confrontation avec les faits, nous allons emprunter une démarche, par éliminations successives des paradoxes, classique dans la recherche.

### 1.2. Approche mécaniste<sup>1</sup> : l'observateur et la relation sujet/objet

La prise de position sur les centrales nucléaires résulte d'une mise en relation d'un sujet : l'homme de la rue, et d'un objet : les centrales nucléaires. Une idée qui vient immédiatement à l'esprit est donc de considérer que cette prise de position est fonction :

- de certaines caractéristiques du sujet : ses connaissances, son savoir, son expérience, certaines dimensions de sa personnalité...
- de certaines caractéristiques de l'objet, que l'on présente souvent, dans le cas des centrales, sous la forme d'une liste d'avantages et d'inconvénients.

---

<sup>1</sup> On a préféré, ici, "mécaniste" à "fonctionnaliste". Voir [Rocher 1968]

Ces caractéristiques sont celles de l'observateur qui fait implicitement l'hypothèse qu'un bilan entre le pour et le contre s'effectue chez l'individu, certains arguments favorables ou défavorables ayant plus de poids que d'autres. C'est au niveau des pondérations qu'il fait intervenir l'expérience, le vécu ou d'autres facteurs liés à l'activité psychique du sujet<sup>1</sup>.

Ce schéma fort simple inspirait la communication dans le domaine du nucléaire dans les années 75. il conduisait à des campagnes d'explication en direction du public reposant sur deux mots d'ordre : informer sur les avantages du nucléaire ; apaiser les craintes.

Mais les campagnes à gros moyens effectuées à cette époque, en Allemagne et en Suède notamment [Salzbourg 1977], n'ont pas donné les résultats escomptés. On avait l'impression que le message envoyé se retournait contre l'émetteur : plus on "informait", plus la remise en cause semblait prendre de l'ampleur. L'inadéquation de ce premier schéma est confirmée par les résultats des sondages. Nous avons vu que, paradoxalement, les mieux renseignés sur l'énergie nucléaire -les milieux intellectuels- étaient, à l'époque, parmi les plus défavorables à l'atome civil ; et que, par contre, les moins informés et les plus impressionnables -les plus âgés- se situaient parmi les plus favorables aux centrales.

*Partir de la relation sujet/objet conduit à un résultat notoirement insuffisant : pourquoi ?*

L'objet "centrales nucléaires" n'est pas un produit familier de consommation courante. C'est un objet médiatique : il est appréhendé avant tout à travers les grands médias ou les discussions que ceux-ci

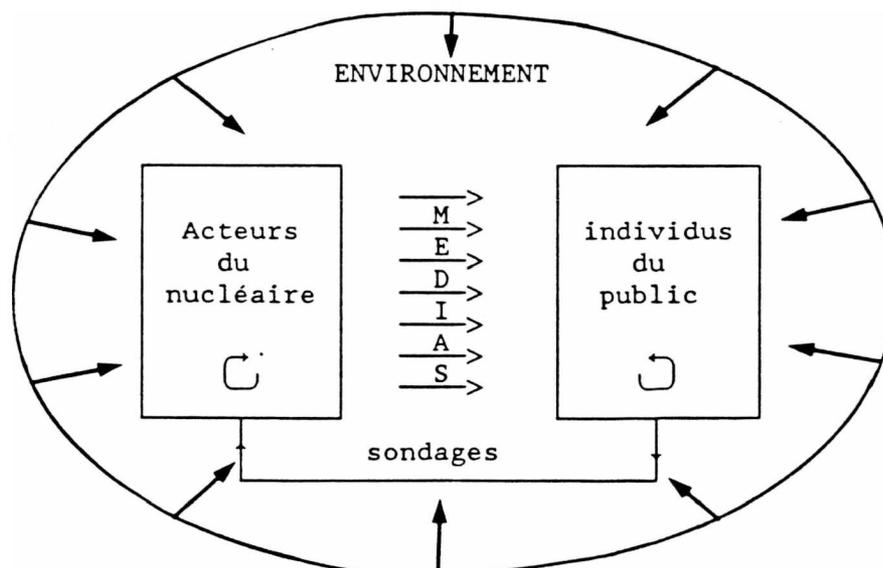
---

<sup>1</sup> On pense ici au modèle de Fishbein-Ajzen [Fishbein-Ajzen 1975] utilisé par H. Otway puis par F. Niehaus de l'Agence Internationale pour l'Energie Atomique de Vienne (voir [Salzbourg 1977]), qui repose sur le schéma croyance → attitude → intention → comportement. Avec ce modèle : tout sujet se forge un certain nombre d'idées, de convictions (croyances) consistant à associer à un objet, avec une intensité (pondération) plus ou moins grande, différents attributs ; l'attitude (l'opinion ici) découle des croyances, c'est à dire, des attributs et de leurs intensités.

provoquent. En ne faisant pas intervenir le processus de communication qui le fait exister auprès du grand public, on est conduit à occulter une partie de ses caractéristiques. Au delà de la technique et du drame (Hiroshima), ce qui relève du symbolique, et qui joue un rôle fondamental dans les représentations, échappe en grande partie à l'analyse. Enfin, en traitant l'individu comme une personne isolée que l'on peut convaincre par les moyens usuels, on oublie les discussions qui ont eu lieu dans le public à propos du programme nucléaire ; à l'évidence, les centrales nucléaires permettent de se situer par rapport aux autres, de se différencier. Donc, il n'est pas suffisant de faire intervenir l'être "psychologique" ; il faut qu'intervienne aussi l'être "social".

### 1.3. Approche constructionniste : la prise en compte du processus de communication

La communication dont l'individu est la cible forge les représentations que l'on a de l'atome civil. Celles-ci se construisent et se transforment, non pas sous l'impulsion des faits et des controverses qu'ils provoquent entre spécialistes et responsables, mais de leurs échos qui nous parviennent à travers les médias. Pour intégrer dans le modèle le processus de communication, il faut savoir circuler des sources jusqu'aux destinataires à travers les canaux de transmission des messages. Aidons nous d'un schéma



La communication dans le domaine nucléaire

Les sources sont les acteurs porte-parole des institutions partie-prenantes dans les enjeux autour desquels se développe, ou s'est développée, la controverse nucléaire. Ici, si on reste au niveau de la France, on retrouve pêle-mêle : l'Etat, le parlement, le gouvernement, les ministères, EDF et le CEA, les groupes d'experts, les partis politiques et les élus, les comités, sociétés ou associations qui soutiennent ou s'opposent au programme nucléaire... et les médias qui, soucieux de leur audience, ne sauraient être considérés comme de simples transmetteurs de l'information. Les enjeux auxquels on fait référence ne sont pas uniquement technologiques ou économiques ; ce sont aussi des enjeux "de pouvoir", chaque organisation oeuvrant, bien entendu, pour défendre ses intérêts et augmenter sa puissance.

Les destinataires, les individus du public, qui ne sont pas véritablement parties-prenantes dans les enjeux, ne peuvent percevoir la réalité qu'à partir des messages émis, tels qu'ils sont transmis par les médias. A leur arrivée, ces messages ne sauraient se réduire aux écrits et déclarations des acteurs ; si on met de côté le commentaire journalistique, ils comprennent aussi bien : "ce qui a été dit", que : "qui l'a dit". La source est partie intégrante du message ; dès qu'elle est perçue de façon négative, ce qui a été dit est ressenti comme suspect.

Mais la communication entre acteurs et individus, qui discutent entre eux du programme nucléaire sans jamais se rencontrer, n'est pas à sens unique. Nous avons vu que la légitimité des acteurs, quand ils se retrouvent sous les projecteurs de l'actualité, peut être remise en cause par l'opinion des gens. L'acteur est alors obligé de tenir compte du public : il va moduler sa position en fonction de l'information sur l'opinion que lui apportent les enquêtes et les sondages. La communication forme donc une boucle où sources et destinataires échangent en permanence leurs rôles ; par un apprentissage continu et réciproque, les acteurs et le public s'ajustent pour tendre vers un équilibre, celui du compromis, qui peut être remis en cause du jour au lendemain. Il suffit d'un événement créé par l'environnement (ex. Tchernobyl) pour que l'harmonie laisse la place à une cacophonie productrice, avec le temps, d'un nouvel ordre.

On voit donc que les messages qui assaillent le public ne constituent pas un simple argumentaire. Prenant en compte les discours et les

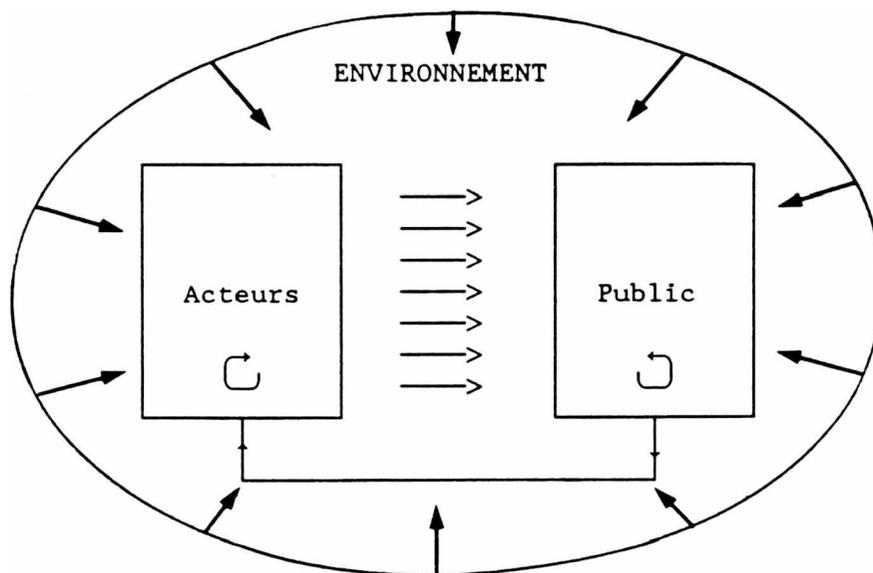
acteurs à la fois, ils forment, mis bout à bout, ce que l'on peut appeler un récit, mieux : une mythologie.

Ce schéma qui intègre le précédent, la relation sujet/objet n'étant pas évacuée mais simplement mise en perspective, est encore insuffisant. Il n'explique en rien pourquoi l'opinion sur les centrales nucléaires se retrouve corrélée avec celle sur la force de frappe, les privatisations ou la publicité... ni pourquoi le facteur politique est aussi déterminant dans les positions prises. Les positions sur les grands conflits qui agitent la société française forment un tout qui paraît bien structuré [Pagès et Coll. 1987] ; comment en tenir compte ?

#### 1.4. Vers une économie des rôles

Parmi les acteurs, nombreux sont ceux qui sont impliqués dans plus d'un conflit. L'Etat, le gouvernement, les ministères, les grandes entreprises, l'église, les sociétés savantes, les partis politiques, les élus, les syndicats, les écologistes... participent, de façon très générale, à cette remise en cause permanente qui propulse "comme des machines à vapeur" nos démocraties, en les transformant, vers les surprises du futur [Charbonnier 1961]. Donc par l'intermédiaire des messages qui font intervenir ce qui est dit et ceux qui le disent, tout se tient : la communication sur un sujet, s'inscrit dans une communication beaucoup plus globale où, dans un espace de discussion, acteurs et individus du public (tous pluridisciplinaires) interagissent. Dans cet espace, les sujets dont on parle sont, bien entendu, en concurrence. Il suffit que l'un émerge avec force, pour ce qui intéressait les médias et l'opinion hier ne l'intéresse plus aujourd'hui. C'est ainsi que Tchernobyl laisse peut-être la place au Sida.

Ainsi nous sommes conduits à juxtaposer les mythologies, à traiter dans leur ensemble les conflits qui, à une époque, mettent en émoi les hommes. Cela conduit à transformer légèrement le schéma précédent. On passe du singulier au global en prenant en compte tous les acteurs qui interviennent dans l'actualité et en élargissant l'espace des sujets de discussion pour intégrer tous les conflits qui comptent.



Le schéma de l'économie des rôles

Le schéma semble simple. On a réduit le phénomène de l'opinion à un système, composé de deux sous-systèmes : "les acteurs" et "le public", ouvert sur un environnement qui tisse les contraintes et produit les faits. En fait, il décrit une réalité fort complexe, difficile à maîtriser à travers des formules. Le système s'auto-régule à deux niveaux : au niveau global par ajustements entre environnement et sous-systèmes des acteurs et du public ; au niveau élémentaire par les ajustements qui s'opèrent entre acteurs d'une part et entre individus d'autre part.

On peut, par analogie, considérer que tout se passe comme dans une économie de marché dont les producteurs seraient les acteurs et les consommateurs les individus du public. Les acteurs en conflit produiraient des rôles que les individus consommeraient pour se différencier. La convergence vers un équilibre serait assurée par trois interactions : la concurrence entre acteurs transparaisant dans le débat public qui les oppose ; le jeu de l'offre et de la demande, où interviennent médias et sondages, permettant d'harmoniser rôle joué par l'acteur et attente du public ; la différenciation où, pour se mettre en valeur par rapport à autrui, l'individu reprend l'un des rôles sur ce marché que constitue la scène des grands médias [Morlat et Pagès 1984].

Voilà où conduit le schéma initial quand on cherche à éliminer les paradoxes par des rajouts successifs. Dans le "schéma de l'économie des rôles" final interviennent, non seulement les caractéristiques premières de l'objet et du sujet, ou les faits et les arguments nouveaux qui sont produits, mais aussi, en général, tout ce qui est à l'origine de la fabrication du sens. L'acteur source du message est pris en compte. Il suffit donc qu'un acteur important change de position pour que les opinions évoluent, et c'est bien ce que l'on a constaté, après mai 1981, suite à l'arrivée au pouvoir de la gauche. Les mouvements contradictoires qui se sont produits depuis 1981, et en particulier ceux décrits dans la figure 1.2, trouvent là une explication.

## BIBLIOGRAPHIE

- Bogardus E. S. (1933). *A social distance scale*, cité dans M.E. Shaw et J.M. Wright, *Scales for the measurement of attitudes*, Mc Graw-Hill, New York, 1967, 409-411.
- Bourdieu P. (1973). *L'opinion publique n'existe pas*. Les temps modernes.
- Charbonnier G. (1981). *Entretiens avec Levi-Strauss*. Plon, coll. 10/18.
- Fishbein M. et I. Ajzen (1975). *Belief, attitude, intention and behaviour : an introduction to theory and research*. Addison-Wesley.
- Fourgous J. M., J. F. Picard et C. Raquenel (1980). *Les Français et l'énergie*. CNRS et EDF, Paris.
- Morlat G. et J.P. Pagès (1984). *Le ciel et la terre. Une approche structuraliste de l'opinion : le baromètre AESOP*, IN : le prix des nouveaux médias, Actes des sixièmes journées internationales, le bulletin de l'IDATE, 17, 50-62.
- Pagès J. P. & Coll. (1987). *Les structures de l'opinion en 1987*. Rapport annuel de l'association pour l'étude des structures de l'opinion Agoramétrie, Paris.
- Rocher G. (1968). *Introduction à la sociologie générale. 2.L'organisation sociale*. Points, Sciences humaines.
- Salzbourg 1977 (1977). *Actes de la conférence internationale sur l'énergie nucléaire et son cycle de combustible*. IAEA, Salzbourg.

## CHAPITRE II

APPROCHE SYSTEMIQUE DE L'ANALYSE FACTORIELLE ET ENQUETES DE  
REFERENCE : DU MODELE STIMULUS-REPONSE AUX ANALYSES FACTORIELLES

### INTRODUCTION

En discutant sur le phénomène de l'opinion on a abouti au schéma de l'économie des rôles qui met en perspective la relation sujet/objet. L'opinion ne dépend pas uniquement des caractéristiques "objectives" du sujet ou de l'objet que l'observateur peut énoncer a priori ; elle dépend, dès qu'il y a conflit, d'autres caractéristiques que l'on ne peut découvrir qu'en analysant ce que l'on a appelé "la différenciation".

On peut considérer que l'objectif des enquêtes sur les structures de l'opinion réalisées par Agoramétrie, est d'analyser les mécanismes de la différenciation. La méthode consiste à rechercher les "dimensions de la différenciation" (structures de l'opinion), en demandant à l'individu (sujet) de se positionner sur un ensemble diversifié de propositions référendum (conflits-objets). On est alors dans la logique d'un modèle : le modèle stimulus-réponse.

C'est à l'étude du modèle stimulus-réponse qu'est consacré principalement ce chapitre où l'analyse factorielle apparaît, non pas comme un ensemble de techniques permettant la description des données, mais comme un moyen pour produire les paramètres d'un modèle qui est satisfaisant dans bien des cas, particulièrement dans les sciences sociales, malgré

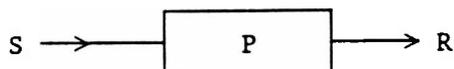
sa simplicité. On retrouvera ici, introduites différemment, certaines des préoccupations des factorialistes de l'époque qui a suivi les premiers travaux de C. Spearman et K. Pearson, et qui conduisent à faire intervenir un univers de variables.

Le modèle stimulus-réponse permet d'introduire la notion d'enquête de référence, c'est à dire d'enquête permettant d'obtenir des dimensions caractéristiques du comportement de l'individu face à une certaine catégorie de situations (les situations permises) : les dimensions de la différenciation dans le cas des enquêtes Agoramétrie sont considérées comme des caractéristiques de l'opinion (comportement) de l'individu dans les grands débats nationaux (catégorie des situations permises).

Les enquêtes de référence définissent un cadre général d'analyse pour l'étude détaillée du comportement d'un groupe d'individus face à une situation ou à un groupe de situations permises particulières, ou encore, face à des situations qui impliquent des comportements qui relèvent, au moins en partie, de la même logique (comportement face au risque, par exemple). Nous verrons qu'il y a deux façons d'envisager les enquêtes de référence selon que l'on suive Spearman ou Pearson, et que cela a son importance pour définir le protocole d'enquête.

### 2.1. En suivant Charles Spearman : approche déterministe des enquêtes de référence

En psychologie quand on s'intéresse aux réactions de l'individu face à une catégorie de situations (les situations permises), on fait en général référence à un schéma du type "S (P) R" :



A ce schéma correspond le modèle :

$$R = f_s(P) \quad (2.1)$$

L'individu est traité comme un système pouvant être dans différents états P ; les réponses R, mesurées en soumettant le système à une

batterie de  $p$  stimuli  $S$  (les situations), sont reliées à l'état  $P$  à travers une fonctionnelle  $f_s$  qui dépend des stimuli. Ce modèle n'est pas valable partout ; il est considéré comme correct dans une certaine "zone d'états". Pour approcher la fonctionnelle  $f_s$ , on peut effectuer une approximation du premier ordre (approximation linéaire) au voisinage d'un individu de référence  $\bar{P}$  pris dans la zone d'états (réponses  $\bar{R}$ ) :

$$R - \bar{R} = \frac{\partial f_s(P)}{\partial P} (P - \bar{P}) + \text{résidu} \quad (2.2)$$

Si dans toute la zone d'états d'intérêt la dérivée  $\partial f_s / \partial P$  peut être considérée comme constante, le modèle (2.2) peut encore s'écrire sous la forme du modèle factoriel [Anastassakos et Gaussens 1983] :

$$\mathbf{x} = U\mathbf{c} + \mathbf{e} \quad (2.3)$$

Dans ce modèle on désigne par :

- $\mathbf{x}$ , le vecteur des réponses centrées (autour de l'individu de référence)  $x^j = R^j - \bar{R}^j$  aux  $p$  stimuli,
- $\mathbf{c}$ , le vecteur des  $k$  paramètres centrés  $c^\ell = P^\ell - \bar{P}^\ell$  permettant de décrire les états du système,
- $U$ , la matrice ( $p \times k$ ) des coefficients  $u_\ell^j$  du modèle (le Jacobien  $\partial f / \partial P$ ) ; ses lignes sont les gradients correspondant aux différents stimuli,
- $\mathbf{e}$ , le vecteur des  $p$  erreurs  $e^j$  faites dans l'approximation  $U\mathbf{c}$  de  $f_s(P)$ .

Quand on réalise une enquête, pour trouver les inconnus  $U$  et  $\mathbf{c}$  du modèle, on a coutume de sélectionner les  $n$  interviewés selon une procédure (la méthode des quotas, par exemple) que l'on assimile à un tirage au hasard dans une population statistique (les français auxquels on s'intéresse). La zone d'états étant assimilée à cette population, on peut placer l'individu de référence au centre de gravité des individus-états sélectionnés. Le modèle (2.3) s'écrit alors, si  $x_i$ ,  $c_i$ ,  $e_i$  sont les  $i$ -èmes réalisations de coordonnées  $x_i^j$ ,  $c_i^\ell$  et  $e_i^j$  des vecteurs  $\mathbf{x}$ ,  $\mathbf{c}$  et  $\mathbf{e}$  respectivement :

$$x_i = U c_i + e_i ; i=1, \dots, n \quad (2.4)$$

Etant ainsi passé d'une approche déterministe à une approche statistique, il reste à préciser ce que sont les paramètres  $U$  et  $\mathbf{c}$ .

Quand on suppose les paramètres  $c^l$  connus, on admet en général que les coefficients  $u_l^j$  ne sont autres que les coefficients de régression des variables  $x^j$  par rapport à l'ensemble des  $k$  paramètres  $c^l$  ; on peut donc imposer dans le modèle (2.3) l'orthogonalité entre les paramètres  $c^l$  et les résidus  $e^j$ . La matrice de variance  $\Sigma$  associée au modèle, se trouve alors décomposée en une somme :

$$\Sigma = \Sigma_1 + \Sigma_2 \quad (2.5)$$

où  $\Sigma_1$  est la matrice de la variance associée à l'approximation  $Uc$  (variance inter) et  $\Sigma_2$  est la matrice de variance associée aux résidus (variance intra).

Mais les paramètres  $c^l$  n'étant pas connus, cette condition n'est pas suffisante pour trouver à la fois les paramètres et les coefficients du modèle ; pour aller plus loin, l'hypothèse supplémentaire à laquelle on a habituellement recours est celle de l'orthogonalité entre les résidus, c'est-à-dire celle de la diagonalité de la matrice  $\Sigma_2$ . Cette hypothèse revient à assimiler ce qui n'est pas expliqué par les paramètres  $c^l$  à du bruit.

C'est bien autour de cette double hypothèse d'orthogonalité qu'ont été conçues toutes les méthodes "Spearmaniennes" d'extraction des paramètres du modèle factoriel [Mulaik 1972]. Pour obtenir  $\hat{\Sigma} = \hat{\Sigma}_1 + \hat{\Sigma}_2$ , estimation de  $\Sigma$  sous la contrainte du modèle factoriel, on rend la matrice  $\hat{\Sigma}_2$  la plus diagonale possible à partir d'une hypothèse sur  $k$  et d'une approximation de  $\Sigma_2$ . On opère par itération en s'appuyant sur différents indices permettant de mesurer la proximité entre la matrice de la variance  $V$  associée aux données  $x_i$  (éq. 2.4) et la matrice  $\Sigma$  associée au modèle (eq. 2.3) :

- avec la méthode MINRES [Harman et Jones 1966], on cherche à rendre minimum la fonction

$$F_1(\Sigma) = 1/2 \text{ trace}(V - \Sigma)^2,$$

- avec la méthode du déterminant maximum<sup>1</sup> [Howe 1955], on cherche à rendre maximum la fonction

$$F_2(\Sigma) = \left| [\text{diag}(V - \Sigma_1)]^{-1/2} (V - \Sigma_1) [\text{diag}(V - \Sigma_1)]^{-1/2} \right|,$$

- etc...

---

<sup>1</sup>. Rappelons que la maximisation de  $F_2(\Sigma)$  conduit à la même solution que la minimisation de la fonction de vraisemblance :

$$F_3(\Sigma) = \log |\Sigma| + \text{trace}(V\Sigma^{-1})$$

Le tirage au hasard effectué, qui peut être assimilé à une perturbation des états (les interviewés) du système (l'individu) dans une zone d'états (la population statistique) où le modèle (2.3) s'applique, permet de garantir que les résultats obtenus par l'analyse factorielle ne dépendent, à travers les interviewés, que de l'univers des individus considérés ; c'est bien pour cela qu'on peut dire que les structures  $U$  et  $c$  dégagées sont des caractéristiques du système (l'individu). Si les conditions d'orthogonalité sont vérifiées, la covariance entre deux stimuli  $x^j$  et  $x^{j'}$  sachant les variables  $c^l$  est nulle :

$$\text{cov}(x^j, x^{j'} / c^1, c^2, \dots, c^k) = 0 \quad (2.6)$$

*L'enquête réalisée sera considérée alors comme une enquête de référence si pour tout autre couple de stimuli correspondant à des situations permises, la relation (2.6) se trouve encore vérifiée : les paramètres d'état  $c^1, c^2, \dots, c^k$  peuvent être alors considérés comme caractéristiques de l'ensemble de toutes les situations permises.*

On retrouve ici, partant de la notion d'enquête de référence, l'un des soucis des psychomètres Spearmaniens du début de ce siècle : l'universalité des résultats que l'on obtient par l'analyse factorielle implique la référence non seulement à un univers d'individus (les états du système), mais aussi à un univers de variables (les stimuli).

#### Lien avec la théorie des images

Comme l'a montré Louis Guttman [Guttman 1953], si l'enquête considérée est une enquête de référence, faire l'analyse factorielle pour trouver les paramètres du modèle (2.3), revient alors à effectuer les  $p$  régressions de chacune des variables  $x^j$  par rapport à l'ensemble  $\{x^l ; l \neq j\}$  des autres variables (théorie des images) ; cette remarque permet d'effectuer l'analyse factorielle sans avoir recours à des itérations [Mulaik 1972] et [Croquette et al., 1984].

On sait en effet que cette "théorie" conduit à décomposer le vecteur  $x$  en deux composantes : la composante "image"  $x_1$  (partie expliquée de variance  $\Sigma_1$ ) et la composante "anti-image"  $x_2$  (résidu de variance  $\Sigma_2$ ).

Si on note  $S^{-2} = \text{diag}(\Sigma^{-1})$  on a :  $x_2 = S^2 \Sigma^{-1} x$  [McDonald 1975]. On en déduit la décomposition suivante de la matrice de la variance :

$$\Sigma = \Sigma_1 + \Sigma_2 - 2S^2(\Sigma^{-1}S^2 - I)$$

Guttman a montré que si le modèle factoriel s'applique sur l'univers de variables considéré, la décomposition image sur tout l'univers coïncide avec la décomposition factorielle. On a "à l'infini" :

$$\Sigma^{-1} \rightarrow S^{-2} ; \quad \Sigma = \Sigma_1 + \Sigma_2.$$

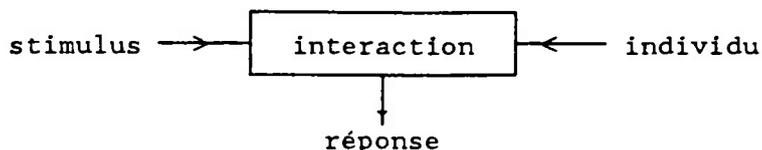
Quand on dispose d'un ensemble de  $p$  variables assez grand et bien diversifié (condition qu'on cherche toujours à réaliser avec l'enquête de référence), la régression d'une variable par rapport aux autres variables (image partielle) revient à la régression par rapport à l'ensemble des variables de tout l'univers (image), régression équivalente, si le modèle factoriel s'applique, à celle qui peut être effectuée par rapport aux composantes factorielles (paramètres  $c^{\ell}$ ).

Avec un grand nombre de variables diversifiées, on peut donc atteindre une représentativité "mathématique" (au sens du modèle factoriel) de l'ensemble des variables auxquelles on fait référence. C'est cette représentativité qui garantit l'universalité des résultats trouvés. Ces derniers ne dépendent que de l'univers infini, non probabilisé, des variables permises. On est ici resté déterministe dans l'appréhension de l'ensemble des variables.

## 2.2. En suivant Karl Pearson : approche statistique des enquêtes de référence

L'approche du modèle stimulus-réponse introduite précédemment peut être qualifiée de "mécaniste" ; on considère que la réponse est produite, à travers l'individu, par des causes (les stimuli) qui préexistent et qui peuvent être décrites par un certain nombre de caractéristiques. C'est ainsi qu'on a abordé la phénomène de l'opinion dans le paragraphe 1.2 en faisant jouer des rôles non symétriques à l'objet (stimulus) et au sujet (l'individu-système).

A cette approche, comme on l'a fait précédemment en 1.3, on peut opposer l'approche "constructionniste" où l'individu et le stimulus sont traités de façon symétrique:



La réponse est ici considérée comme résultant, non pas des caractéristiques de l'individu et du stimulus pris isolément, mais de ce nouvel objet qu'ils forment dans leur interaction. On fait l'hypothèse que la signification pour l'individu du stimulus se construit au fur et à mesure d'un apprentissage au cours duquel l'individu se transforme en donnant son sens au stimulus : l'un agit donc sur l'autre, et réciproquement.

Le système d'enquêtes d'Agoramétrie fournit un bel exemple de l'approche constructionniste : ici, les individus prennent position dans les grands conflits nationaux en fonction de la signification que ces conflits ont pour eux, significations, qui ne peuvent être découvertes qu'en analysant leur prise de position. On abandonne la dénotation (caractéristiques "objectives" du stimulus) et le statut "objectif" (caractéristiques de l'individu), qui sont les concepts auxquels a recours habituellement le psycho-sociologue, pour deux nouvelles notions qui se renvoient l'une à l'autre :

- *la connotation* : elle traduit la véritable signification qui est accordée par l'individu social au stimulus-conflit.
- *l'attitude de différenciation* : elle est la prédisposition à agir que provoque chez l'individu le conflit. Cette prédisposition consiste à utiliser l'occasion pour exprimer sa différence, c'est à dire son statut, tel qu'on le perçoit ou tel qu'on veut qu'il soit.

On est dans un cercle vicieux qui ne peut être rompu qu'en introduisant la notion d'interaction. Dans l'approche précédente le système était l'individu ; dans l'approche constructionniste le système est l'interaction elle même. C'est en analysant cette interaction que l'on pourra produire des caractéristiques de l'individu et des situations qui devront se correspondre "par dualité".

Le modèle auquel on fait référence, s'écrit ici :

$$R = f(S,P) \quad (2.7)$$

où, R désignant toujours la réaction de l'individu, S et P sont les ensembles de paramètres qui permettent de traduire dans leur interaction le stimulus (S) et la réponse (P) ; et où les mécanismes qui conduisent à la prise de position sont représentés par la fonctionnelle f.

Considérer l'interaction comme système, c'est assimiler les paramètres qui décrivent le stimulus et l'individu à des paramètres d'état. De même que précédemment, on peut se fixer une certaine zone d'états où une approximation linéaire du premier ordre est considérée comme satisfaisante.

Si on emprunte une démarche statistique, le système est alors perturbé dans cette zone à l'aide d'un double tirage au hasard : le premier consiste toujours à tirer un échantillon d'interviewés d'une population d'individus particulière ; le second conduit, de façon similaire, à trouver les stimuli parmi les situations permises en assimilant l'ensemble de ces situations à une population statistique.

Cette double perturbation conduit, si on se contente d'une approximation du premier ordre, à un modèle linéaire. Si on note comme précédemment  $c^1, c^2, \dots, c^k$  les paramètres caractérisant l'état du système, on peut écrire :

$$x_i^j = \sum_{\ell=1}^k c_{i\ell}^j u_\ell^j + e_i^j \quad ; i=1, \dots, n ; j=1, \dots, p \quad (2.8)$$

Si on fixe les p stimuli on retrouve le modèle (2.4) :

$$x_i = U c_i + e_i \quad ; i=1, \dots, n \quad (2.9)$$

Si on fixe les n individus, les équations (2.8) s'écrivent :

$$x^j = C' u^j + e^j \quad ; j=1, \dots, p \quad (2.10)$$

où :  $C(k \times n)$  désigne la matrice des coefficients  $c_i^\ell$  dont les colonnes sont les vecteurs  $c_i$  ;  $x^j$  et  $e^j$  les n-vecteurs de coordonnées  $x_i^j$  et  $e_i^j$  ; et  $u^j$  la j-ème colonne de la matrice U.

Si on suit la même démarche qu'en 2.1., on doit admettre que dans l'équation (2.9) les coefficients  $u_\ell^j$  sont les coefficients de régression des variables  $x_i^j$  par rapport aux paramètres  $c^\ell$ . Le fait de traiter de façon symétrique les individus et les variables conduit à estimer les

paramètres du modèle (2.10) de la même manière que dans le cas du modèle (2.9) : ce qui est vrai pour (2.9) doit être vrai encore quand on transpose (dualité), pour (2.10). Dans cette dernière équation les coefficients  $c_i^l$  doivent donc être les coordonnées de la projection (coefficients de "régression") des individus  $x_i$  sur le sous-espace engendré par les vecteurs  $u_l$ . Cette double contrainte, que doit respecter (2.8), est caractéristique des solutions qu'offre l'analyse en composantes principales.

Considérons les individus  $x_i$  comme des points de  $E=R^p$  et les variables  $x_j$  comme des points de  $F=R^n$ . Si on munit  $E$  de la métrique euclidienne  $M$  et  $F$  de la métrique "des poids"  $D_p$  -  $X$  désignant le  $(p \times n)$  tableau centré des données  $x_i^j$  - au triplet  $(X, M, D_p)$  est associé le schéma de dualité [Cailliez et Pagès 1976] :

$$\begin{array}{ccc}
 E=R^p & \xleftarrow{X} & F^* \\
 \begin{array}{c} \uparrow \\ M \\ \downarrow \\ V \end{array} & & \begin{array}{c} \uparrow \\ W \\ \downarrow \\ D_p \end{array} \\
 E^* & \xrightarrow{X'} & F=R^n
 \end{array}$$

Dans ce schéma :

$$V = X D_p X'$$

$$W = X' M X$$

Les conditions d'orthogonalité qui viennent compléter les équations (2.9) et (2.10), conduisent aux deux décompositions en somme directe duales :

$$\begin{array}{lcl}
 E & = & E_1 + E_2 \quad ; \quad X'(E^*) = F_1 + F_2 \\
 \uparrow & & \uparrow \quad \quad \quad \uparrow \\
 x_i & = & U c_i + e_i \quad ; \quad x_j = C' u_j + e_j
 \end{array}$$

$$\text{avec :} \quad E_1 \perp_M E_2 \quad \text{et} \quad F_1 \perp_{D_p} F_2.$$

On en déduit :

$$E_1 \perp_M E_2 \quad \text{et} \quad E_1 \perp_{MVM} E_2.$$

Cette double condition d'orthogonalité pour les espaces  $E_1$  et  $E_2$  n'est vérifiée que si [Cailliez et Pagès 1976, p.276] les espaces sont engendrés par deux sous-ensembles complémentaires de l'ensemble des vecteurs propres de VM. Pour que les résidus soient minimaux, on retiendra donc pour  $E_1$  le sous-espace engendré par les  $k$  premiers axes principaux (vecteurs  $u_p$ ) du triplet  $(X, M, D_p)$  ;  $F_1$  est alors engendré par les  $k$  premières composantes principales (vecteurs  $c^k$ ).

En général, en analyse en composantes principales, on prend pour métrique M la métrique  $D_{1/\sigma^2}$  (ACP sur données centrées réduites) ; mais d'autres choix sont possibles comme celui de la "métrique de Jöreskog" :

$$M = \text{diag}(V^{-1}) = S^{-2}.$$

Dans ce cas l'analyse en composantes principales du triplet  $(X, M, D_p)$  revient à une analyse factorielle au sens de Spearman effectuée par le biais de la théorie des images [Croquette, Kobilinski et al. 1984].

Remarque : si on veut traiter véritablement de façon symétrique les individus et les variables, la métrique M devrait être prise diagonale. Les éléments diagonaux seraient alors les poids accordés -et cela en toute rigueur devrait être fait a priori- aux variables dans le calcul des distances entre individus.

Voulant traiter dans l'approche "constructionniste" les individus et les variables de façon symétrique, on a été conduit à perturber le système auquel a été assimilée l'interaction à l'aide d'un double tirage au hasard. Cette procédure permet d'obtenir une enquête qui peut être considérée, dans un sens différent que précédemment, comme une enquête de référence, le double tirage conduisant, par l'analyse en composantes principales, à des résultats qui sont indépendants des individus interrogés et des variables retenues tout à la fois.

Avec Spearman (analyse factorielle en facteurs communs et spécifiques (AFS)), la notion d'enquête de référence repose sur un modèle mathématique caractérisant l'univers des variables. La qualité des résultats est dépendante du mode de sélection (des variables) qui doit conduire à des variables diversifiées en assez grand nombre. Avec Pearson (analyse en composantes principales (ACP)) la notion d'enquête de référence ne fait intervenir aucune hypothèse sur l'univers des

variables. La qualité des résultats dépend ici de la capacité que l'on a à tirer au hasard dans cet univers.

Ainsi, on a glissé d'une approche déterministe (de l'univers des variables), où la difficulté principale réside dans le "modèle", à une approche statistique où la difficulté principale réside dans le "tirage au hasard". Il faut remarquer que, quel que soit le type d'analyse globale effectuée (ACP avec métrique quelconque, AFS, Classification automatique, etc), l'approche statistique garantit des résultats universels.

### 2.3. Réalisation d'une enquête de référence

Dans l'élaboration d'une enquête de référence deux raisons nous font préférer l'approche statistique par l'ACP :

- elle est cohérente avec la démarche constructionniste qui est bien celle qui nous paraît convenir quand on aborde l'étude des opinions. On est loin ici des préoccupations des psychomètres du début de ce siècle qui, centrés sur l'individu, cherchaient à bâtir des batteries de tests pour mesurer l'intelligence ou décrire la personnalité. On cherche ici aussi bien à approfondir ce qu'est l'univers des objets (quelle est la signification des stimuli-questions ?) que l'univers des sujets (qu'expriment les individus ?).
- elle est bien dans l'esprit d'une approche système où on circule continuellement entre l'élément (l'individu, l'acteur), le sous-système (les individus, les acteurs) et le système (la relation acteurs-public). Pour comprendre le phénomène, on ne peut pas se contenter de rester au niveau global, c'est à dire au niveau des deux univers (individus et variables) dont les dimensions factorielles sont les caractéristiques ; Il faut aller examiner dans le détail ces éléments particuliers que sont les individus et les variables sélectionnés. Or en ACP, grâce à la dualité, on n'éprouve pas de difficulté pour calculer les coordonnées des individus sur les axes factoriels, comme c'est le cas en AFS (problème de l'estimation des scores). Et c'est ainsi que l'on peut exploiter au mieux (par la procédure des points supplémentaires) les informations dont on dispose aussi bien du côté des individus que du côté des variables.

On a toujours une idée suffisamment précise de l'univers des individus ; il peut être, par exemple, l'ensemble de français majeurs vivant en métropole. Diverses procédures ont été développées pour effectuer le tirage au hasard dans cet univers, celle reposant sur la méthode des strates et des quotas étant la plus connue. Notons que ces procédures sont plus complexes qu'il n'y paraît ; en plus de la méthode, elles font intervenir bien d'autres paramètres : comment caractériser, par exemple, le réseau d'enquêteurs sur lequel on s'appuie pour sélectionner les interviewés ? comment caractériser la manière d'aborder les gens ? Du côté des individus, on peut donc considérer qu'adopter une procédure revient, en gros, à choisir un institut de sondage.

Du côté des variables, on se heurte d'abord à un problème de définition : quel est véritablement l'univers auquel on fait référence ? quand on parle de conflits, de quels conflits parle-t-on ? quand on aborde les mécanismes de la perception des risques, à quelles situations pense-t-on ? bref, que recouvre le concept que l'on fait intervenir quand on parle des variables ? Et, même si ce concept est clair, puisqu'il faut probabiliser : quelle distribution de masses considérer sur l'univers ? Compte tenu des milliers de façons de libeller une question, on est certainement ici plus proche d'un continuum que du côté des individus où souvent une distribution uniforme des masses s'impose (l'enquête d'opinion repose en général, comme le vote, sur l'idée du suffrage universel).

Mais en réalité qu'importe que l'on ne sache pas tirer dans une population statistique ! L'essentiel est que l'on puisse assimiler l'ensemble de variables sélectionnées à un échantillon tiré au hasard. Et c'est bien le cas, si on se donne une procédure de production de variables qui, tout en étant erratique, est par ailleurs suffisamment précise pour respecter certaines règles. C'est la procédure elle-même qui est caractéristique à la fois de l'univers et de la distribution des masses.

Pour illustrer ces procédures de production de variables assimilées à des tirages au hasard, nous allons indiquer comment on procède à Agoramétrie [Pagès et coll. 1987]. Pour construire les questionnaires permettant d'extraire les "structures de l'opinion" (enquête de référence au sens de l'ACP), on opère en trois étapes :

- sélection d'un grand ensemble de conflits diversifiés à l'aide d'une méthode fixée une fois pour toutes ; cet ensemble est assimilé à l'univers des conflits.

- extraction par un groupe d'experts d'un échantillon représentatif de cet univers.

- construction, autour des conflits répertoriés, de propositions référendum sans ambiguïté.

#### ■ Construction d'un univers de conflits

Une analyse systématique des grands titres de la presse nationale et régionale (des quotidiens : L'Equipe, Le Figaro, France-Soir, L'Humanité, Libération, Le Monde, Ouest-France... et des magazines : Actuel, Canard Enchaîné, Elle, L'Express, Le Nouvel Observateur...) est effectuée un mois avant le déroulement de l'enquête.

Les cinq à six cent conflits répertoriés sont rangés suivant une nomenclature fixée une fois pour toutes ; en voici les rubriques : politique française, politique internationale, affaires françaises bizarres, catastrophes et problèmes nationaux et internationaux, économie, sécurité, justice, famille, travail et moyens de lutte, médical, urbanisme et logement, énergie, armée, enseignement, religion, agriculture, environnement et sauvegarde de la nature, transports, mœurs, problèmes de société, sciences, information et médias, le 3ème âge, consommation et alimentation, qualité et mode de vie, spectacle et littérature, sports et loisirs, mode et beauté.

Pour avoir une idée de ce que peut être une rubrique donnons le contenu de la rubrique "Problèmes de Société" en 1987 : le banditisme, la criminalité, l'immigration, le racisme, le code de la nationalité, la drogue, la violence, les conséquences du sida, la délinquance juvénile, les suicides, les vols, le désarroi de certains jeunes.

Ce qu'on appelle conflit dépasse évidemment ce qu'on peut recenser à travers la presse ; l'hypothèse qui est faite est qu'en opérant ainsi on obtient un ensemble de conflits explicites aussi divers que l'univers auquel il est fait implicitement référence.

## ■ Echantillonnage

Un groupe d'experts, bien au fait des enquêtes et aux pôles d'intérêt diversifiés, se mettent d'accord pour retenir, après des discussions, parfois fort vives, plus de soixante conflits dans l'ensemble précédent considéré comme un univers. Voici certains des conflits qui ont été sélectionnés dans toutes les enquêtes depuis 1977 : l'avortement, la force de frappe, l'immigration, la télévision, la croissance.

On ne peut pas parler ici explicitement de tirage au hasard ; il y a simplement production d'un certain nombre de conflits suivant une procédure qui permet à chacun des experts d'exprimer ses convictions : tout se passe comme si chaque conflit était muni d'une probabilité subjective traduisant l'ensemble de toutes ces convictions.

## ■ Construction des propositions référendum

Les propositions, sur lesquelles les interviewés se prononcent grâce à une échelle en 5 paliers, sont bâties autour des conflits précédents. Par exemple, si on reprend les conflits cités précédemment, les formulations correspondantes sont :

- la libéralisation de l'avortement est une bonne chose
- la force de frappe est indispensable
- il y a trop de travailleurs immigrés
- ils nous prennent pour des abrutis à la télévision
- on doit tout faire pour la croissance économique

La phrase construite ne comprend, en général, qu'un seul terme fort faisant référence au conflit ; pour qu'elle soit sans ambiguïté, il importe de ne retenir qu'un seul "signifié". La proposition doit de plus être assez proche des formules qui sont utilisées pour débattre du conflit dans le public.

## BIBLIOGRAPHIE

- Anastassakos I. & E. Gaussens (1984). *Approche systémique de l'analyse factorielle. Les problèmes posés par les deux univers*. Statistique et Analyse des Données, 8, 1-15.
- Cailliez F. & J.P. Pagès (1976). *Introduction à l'analyse des données*. Paris : SMASH.
- Croquette A., A. Kobilinsky, J. P. Pagès & Y. Schektman (1983). *Optimalité en analyse factorielle*. Cahiers du B.U.R.O. n°43-44, 3-41.
- Guttman, L. (1953). *Image theory for the structure of quantitative variates*. Psychometrika, 18, 277-296.
- Harman, H. H. & W. H. Jones (1966). *Factor analysis by minimizing residuals*. Psychometrika, 31, 351-368.
- Howe, W. G. (1955). *Some contributions to factor analysis*. Report No. ONRL-1919, Oak-Ridge, Tenn.: Oak-Ridge National Laboratory.
- McDonald, R. P. (1975). *Descriptive axioms for common factor theory, image theory and component theory*. Psychometrika, 40, 137-152.
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York : McGraw-Hill.
- Pagès J. P. & Coll. (1987). *Les structures de l'opinion en 1987*. Paris : édité par l'association pour l'étude des structures de l'opinion Agoramétrie.

## CHAPITRE III

### ENQUETES DE REFERENCE ET COMMUNICATION ENTRE ENQUETES

#### INTRODUCTION

La réalisation d'une enquête de référence, où on cherche à être représentatif du point de vue des individus et des variables tout à la fois, nécessite le recours à des procédures assez lourdes et coûteuses comme celle exposée en 2.3. On comprend alors l'intérêt d'une exploitation optimale des résultats fournis par les analyses statistiques de cette enquête, résultats qui sont censés être caractéristiques du phénomène étudié et qui peuvent, par conséquent, constituer un cadre pour l'analyse de phénomènes particuliers. Pour exploiter ce cadre, lors de la réalisation d'une étude dont le domaine d'intérêt relève, au moins en partie, du domaine de l'enquête de référence, on tâchera à inclure dans le protocole de l'enquête particulière effectuée des éléments permettant la mise en perspective.

Si la comparaison des tris à plat et des tris croisés entre enquêtes ne fait pas appel à des techniques statistiques particulières, il n'en est pas de même pour les résultats des analyses multidimensionnelles. Nous verrons que la conception de la variable en tant qu'"objet", faisant référence à une population (voir section 2.2), soulève bien des questions quand on essaye de comparer des espaces factoriels issus de populations d'individus différentes enquêtées au même moment (synchronie). Une situation analogue se présente lorsqu'on compare les indicateurs factoriels relatifs à des enquêtes, sur une même population d'individus, ayant lieu à des moments différents (diachronie).

La méthodologie développée dans ce chapitre concerne, en premier lieu, la situation où le traitement statistique approfondi des données et l'interprétation des résultats d'une enquête de référence précède les analyses de certaines enquêtes particulières ; les dernières sont précisément conçues et réalisées de façon à pouvoir être analysées à la lumière des résultats d'une telle enquête de référence.

L'objectif est de faire passer (communiquer) l'information contenue dans les indicateurs factoriels de l'enquête de référence (notée enquête A) dans une enquête particulière (notée enquête B), la communication pouvant éventuellement s'effectuer dans le sens inverse. Nous avons vu dans le chapitre 2 que ces indicateurs (les dimensions de la différenciation) sont des dimensions caractéristiques du comportement de l'individu face à une catégorie de situations ; ayant été laborieusement interprétés dans A, on veut qu'ils servent de véritable référence (indicateurs de référence) dans l'analyse de B.

Pour rendre la communication possible, on insère dans le questionnaire de B des variables de A soigneusement sélectionnées qu'on appellera variables de la passerelle. Le passage de l'information ne peut alors s'effectuer que si on est capable de retrouver une structure équivalente à celle de A dans un espace factoriel significatif de B auquel peuvent contribuer les variables de la passerelle. Quelle que soit la technique utilisée, ces variables permettront le recalage des dimensions factorielles de B sur celles de A.

Ici, il faut attirer l'attention sur le fait que deux variables communes, c'est à dire ayant le même libellé, ne sont pas forcément équivalentes dans A et B. La signification des variables, qui n'est appréhendée qu'a posteriori en analysant les réactions (différenciation) provoquées chez les individus interrogés, peut varier d'une enquête à l'autre<sup>1</sup>. Ce type de variation est à l'origine de perturbations dans les matrices de variance entre les variables communes (différences impor-

---

<sup>1</sup> Quand on fait des enquêtes ayant lieu à des époques différentes (diachronie), une proposition, par exemple "le gouvernement manque d'efficacité", peut prendre des significations complètement différentes. Il en est de même lorsqu'à une même époque (synchronie) on enquête dans des populations distinctes : la proposition "les étudiants vivent en parasites de la société" n'est pas perçue par une population d'étudiants de la même façon que par une population de retraités.

tantes entre lignes et colonnes homologues) qui conduisent à des aberrations si on fait communiquer sans précaution les enquêtes.

On commence dans ce chapitre par introduire un ensemble de méthodes auxquelles on peut penser pour traiter du problème, tel qu'il a été posé, de communication entre enquêtes. On jugera de l'adéquation des méthodes en tenant compte des remarques précédentes et du fait que la procédure de communication doit permettre d'exploiter l'information apportée à la fois par les variables et par les individus (par l'intermédiaire des variables qualitatives illustratives). Notre propre méthodologie sera alors introduite ; nous verrons qu'elle permettra de pallier à certaines insuffisances.

### 3.1. Présentation critique de trois méthodes

Rechercher une structure dans l'espace factoriel de B, sur la base de la connaissance acquise en A, relève de ce que les anglo-saxons appellent "analyse factorielle confirmatoire". Se situent dans ce cadre les travaux, bien connus, de Jöreskog qu'on passe d'abord en revue. Ici, pour mieux respecter la démarche statistique inférentielle empruntée, on a gardé les notations classiques :  $\Sigma$  pour la matrice de variance théorique, S pour la matrice de variance observée... Dans les limites de la validité des hypothèses faites, la méthodologie exposée permettra l'analyse fine des écarts entre matrices de variance ainsi que la mise à l'épreuve de toute hypothèse structurelle sur un espace factoriel.

Ne faisant plus appel au modèle probabiliste, on reprend ensuite les notations de l'Analyse de Données (V- $\Sigma$ -S) pour présenter "L'Analyse Conjointe de Plusieurs Tableaux de Données" ; il s'agit d'une méthode, bien connue en France, pour la comparaison, dans un cadre très général, des espaces factoriels.

On introduit, enfin, "L'Analyse de Redondance", une technique complémentaire aux précédentes relativement au problème de communication posé. Ici, l'objectif n'est pas l'étude des structures factorielles à travers les enquêtes, mais l'analyse, à l'intérieur d'une enquête

particulière, des variables de structure inconnue sous la contrainte imposée par la structure des variables de référence (variables de la passerelle).

### 3.1.1. L'approche de Jöreskog

L'approche initiale de Jöreskog [Jöreskog 1969 et 1971] qu'il appelle "analyse factorielle confirmatoire" se situe dans le cadre du modèle factoriel classique (modèle en facteurs communs et spécifiques). Sous l'hypothèse d'un tel modèle, la matrice de la variance  $\Sigma_i(p_i \times p_i)$  d'une population  $i$  admet la décomposition :

$$\Sigma_i = \Lambda_i \Phi_i \Lambda_i' + \Psi_i^2 \quad ; \quad \Psi_i^2 \text{ diagonale} \quad (3.1)$$

où,

$\Lambda_i(p_i \times k_i)$  est la matrice des coefficients de régression ("saturations") des variables par rapport aux composantes factorielles ("facteurs").

$\Phi_i(k_i \times k_i)$  est la matrice des variances-covariances des facteurs,  
 $\Psi_i^2(p_i \times p_i)$  est la matrice diagonale des variances résiduelles ("unicités").

Soit  $n_i$  la taille de l'échantillon disponible ( $n_i$  observations sur  $p_i$  variables). Notons  $X_i(p_i \times n_i)$  le tableau de données centré correspondant et  $S_i(p_i \times p_i)$  la matrice de variance observée :  $S_i = \frac{1}{n_i} X_i X_i'$ . Sous l'hypothèse de multinormalité des observations, le logarithme de la vraisemblance pour l'échantillon étudié s'écrit alors à une fonction des observations près :

$$\text{Log } L_i = \frac{-1}{2} n_i ( \log |\Sigma_i| + \text{tr}(S_i \Sigma_i^{-1}) - \log |S_i| - p_i ) \quad (3.2)$$

Si on dispose de  $m$  échantillons indépendants provenant de  $m$  populations, alors le logarithme de la vraisemblance totale  $\text{Log } L$  est égal à  $\text{Log } L = \sum_{i=1}^m \text{Log } L_i$ .

Sous l'hypothèse d'une décomposition factorielle pour chaque population, on obtient des estimations simultanées  $\hat{\Lambda}_i, \hat{\Phi}_i, \hat{\Psi}_i^2$  ;  $i=1, \dots, m$ , en minimisant  $F = -\text{Log } L$  par rapport à  $\Lambda_i, \Phi_i, \Psi_i^2$ . Cette minimisation s'effectue en tenant compte du fait que certains éléments des matrices à

estimer peuvent être fixés à des valeurs préspecifiées ou contraints à être égaux entre eux ; ce sont des contraintes de ce type qui permettront de tester, en ayant recours à la statistique du rapport de vraisemblance, toute hypothèse d'identité structurelle entre les populations. Car, si l'hypothèse factorielle (sous contraintes) est vraie, deux fois la valeur de F au minimum est égal à la valeur prise par la statistique, qui est donc distribuée comme un khi-2 à  $\sum_{i=1}^m \frac{1}{2} p_i (p_i + 1) - t$  degrés de liberté (d.d.l.), où t est le nombre total de paramètres indépendants estimés dans le modèle<sup>1</sup>.

Pour mettre en évidence toute caractéristique structurelle pouvant être considérée comme invariante entre les populations, Jöreskog [Jöreskog 1971] propose une démarche exploratoire inférentielle qui consiste en une succession d'analyses : il s'agit de restreindre à chaque étape le modèle factoriel étudié en introduisant des spécifications supplémentaires sur les paramètres qui tiennent compte des verdicts des tests effectués.

L'"analyse des structures de covariance" peut être regardée comme une généralisation de l'analyse factorielle confirmatoire. Il s'agit là d'une approche très générale pour l'étude des relations structurelles entre variables.

Selon Jöreskog [Jöreskog 1978] une structure de covariance peut être, de façon très générale, définie en spécifiant que les éléments  $\sigma_{ij}$  de la matrice de variance  $\Sigma(p \times p)$  d'une population étudiée, sont

---

<sup>1</sup>Donnons ici des exemples :

Soit  $\varphi_{uv}^i$  et  $\lambda_{jl}^i$  les éléments généraux des matrices respectivement  $\Phi_i$  et  $\Lambda_i$ . L'orthogonalité entre les facteurs de la i-ème population est obtenue en effectuant la minimisation de F ayant imposé  $\varphi_{uv}^i = 0$ ,  $u \neq v$  ; et éventuellement  $\varphi_{uu}^i = 1$  (normalisation).

Si la variable  $x^a$  est commune aux populations  $i_1$  et  $i_2$ , on testera souvent l'hypothèse dite "d'invariance" en posant :  $\lambda_{al}^{i_1} = \lambda_{al}^{i_2}$  ; ou encore  $\lambda_{al}^{i_1} = \lambda_{al}^{i_2} = c_{al}$  = valeur spécifiée, si les valeurs des saturations  $c_{al}$  sont supposées connues (comme, par exemple, en présence d'une enquête de référence).

des fonctions des paramètres  $\theta_1, \theta_2, \dots, \theta_t$ ,<sup>1</sup> i.e.  $\Sigma = \Sigma(\theta)$ . (De façon analogue on peut définir une structure de corrélation P, en posant  $\Sigma = D_\sigma P(\theta) D_\sigma$  ou simplement  $P=P(\theta)$ ). Pour ajuster  $\Sigma$  à la matrice de la variance S observée sur un échantillon de taille n, on utilise le plus souvent l'une des trois méthodes suivantes :

■ La méthode des moindres carrés non pondérés (M.C.N.P.), où on minimise :

$$U = 1/2 \text{ trace}(S - \Sigma)^2 ,$$

■ La méthode des moindres carrés généralisés (M.C.G.), où on minimise :

$$G = 1/2 \text{ trace}(I - S^{-1}\Sigma)^2 ,$$

■ La méthode du maximum de vraisemblance (M.V.), où on minimise :

$$M = \text{trace}(\Sigma^{-1}S) + \log|\Sigma| - \log|S| - p.$$

La minimisation se fait par rapport à  $\theta$ , certains parmi les  $\theta_i$  pouvant être soumis à des contraintes.

Les trois fonctions (U, G, M) peuvent être minimisées par le même algorithme numérique (algorithme de Fletcher-Powell, Gauss-Newton, Fisher's scoring ou autre). Le recours à la méthode M.V. trouve une justification statistique dans le cas de la multinormalité ; il en est de même pour la méthode M.C.G.. Au minimum (2/n) fois la matrice des dérivées secondes de M ou de G par rapport à  $\theta$  fournit une estimation de la matrice de covariance asymptotique de l'estimateur  $\hat{\theta}$  de  $\theta$  ; ses éléments diagonaux permettent donc la construction d'intervalles de confiance pour les  $\theta_i$ . De plus au minimum (n/2) fois M (ou G) coïncide avec la statistique du rapport de vraisemblance qui est distribuée, dans des échantillons de taille importante, comme un khi-2 à  $1/2p(p+1)-t$  d.d.l. et peut alors être utilisée comme un indice de la qualité d'ajustement global. Pour une discussion sur la puissance de cette statistique qui joue un rôle important dans ce modèle on consultera [Satorra et Saris 1985] et [Matsueda et Bielby 1986].

---

<sup>1</sup>Dans le cas particulier où  $\Sigma$  accepte la décomposition factorielle :  $\Sigma = \Lambda \Phi \Lambda' + \Psi^2$ , les paramètres  $\theta_i$  sont les éléments  $\lambda_{j\ell}$ ,  $\varphi_{uw}$  et  $\psi_{jj}$  des matrices  $\Lambda$ ,  $\Phi$  et  $\Psi^2$ . Ici,  $\sigma_{ij}(\theta) = \sum_{\ell=1}^k \lambda_{j\ell} \sum_{v=1}^k \lambda_{iv} \varphi_{v\ell} + \psi_{ij}$ , où k est le nombre des facteurs.

On peut tester différentes hypothèses structurelles (emboîtées) dans ce modèle en tenant compte du fait que si  $F_t$  est le minimum de  $M$  (ou de  $G$ ) dans le modèle général ( $t$  paramètres à estimer) et  $F_v$  le minimum obtenu sous une hypothèse structurelle (en imposant  $t-v$  contraintes supplémentaires, ce qui laisse donc  $v \leq t$  paramètres indépendants à estimer), alors  $b = (n/2)(F_v - F_t)$  est distribué comme un  $\chi^2$  à  $t-v$  d.d.l..

Jöreskog préconise l'utilisation de  $b$ , non pas comme une statistique, mais plutôt comme un indice d'ajustement, à cause de la sensibilité du  $\chi^2$  aux différentes hypothèses comme la linéarité, l'additivité, la normalité, etc. Le problème que l'on se pose en réalité, n'est pas statistique : il ne s'agit pas de tester une certaine hypothèse sur  $\Sigma$  (qui à priori peut être considérée comme fausse), mais d'ajuster différents modèles, faisant intervenir un plus ou moins grand nombre de paramètres, et de faire un choix parmi ces modèles. On s'intéresse beaucoup plus à la variation du  $\chi^2$  d'un modèle à l'autre, et cela en tenant compte des degrés de liberté, qu'à la valeur absolue du  $\chi^2$ .

Ajoutons ici que dans le cas de la méthode M.V., par exemple,  $b = n/2(F_v - F_t)$  peut se réécrire :

$$b = (n/2) \{ \log[ |\hat{\Sigma}_v| / |\hat{\Sigma}_t| ] + \text{trace}[ S(\hat{\Sigma}_v^{-1} - \hat{\Sigma}_t^{-1}) ] \},$$

où  $\hat{\Sigma}_t$  et  $\hat{\Sigma}_v$  sont les estimations de la matrice de covariance sous respectivement  $t$  et  $v$  contraintes. Pour un écart fixe entre  $\hat{\Sigma}_v$  et  $\hat{\Sigma}_t$ , la statistique  $b$  croît alors proportionnellement à la taille de l'échantillon. Pour compenser l'influence de  $n$  sur la qualité d'ajustement, plusieurs auteurs ont proposé des indices d'ajustement non directement liés à  $n$  (voir par exemple [Jöreskog et al. 1984]). Notons toutefois que l'approche qui consiste à s'affranchir de la taille de l'échantillon dans l'expression de l'indice d'ajustement est très critiquable d'un point de vue statistique (voir, par exemple, [Matsueda et al. 1986]).

L'approche globale précédemment décrite, développée en particulier par Jöreskog (pour le cas d'une analyse simultanée sur différentes populations, on consultera [Lee 1982]), a été rendue opérationnelle grâce à différents progiciels dont LISREL [Jöreskog et Sörbom 1984] et COSAN [McDonald 1980]. On trouvera une discussion sur les liens qu'entretiennent les deux modèles ainsi qu'avec le modèle RAM dans [McArdle

et McDonald 1984] et [McDonald 1986]. Dans COSAN notamment, toute décomposition de la forme :

$$\Sigma = \left( \prod_{j=1}^m F_j \right) P \left( \prod_{j=1}^m F'_j \right) , \quad (3.3)$$

où  $P$  est symétrique, peut être considérée ; les éléments des  $F_j$  (ou de  $F_j^{-1}$  éventuellement) et de  $P$  sont des fonctions des paramètres  $\theta_1, \theta_2, \dots, \theta_t$ , sur lesquels tout un ensemble de contraintes (linéaires ou non) peut par ailleurs être imposé. Avec ce modèle, cas particulier du modèle général  $\Sigma = \Sigma(\theta)$ , on arrive à traiter la quasi-totalité des applications connues relevant de l'analyse des structures de covariance.

Pourquoi va-t-on en partie se démarquer de l'approche de Jöreskog ?

- d'abord, parce que les procédures proposées perdent de leur intérêt, dès que l'on s'écarte de la multinormalité ; c'est bien le cas dans les enquêtes psycho-sociales.

- puis, parce que les calculs (minimisation par des méthodes numériques) sont très coûteux dès que le nombre de paramètres à estimer (sous contraintes) est grand. Rappelons que dans nos enquêtes on manipule une centaine de variables, ce qui entraînerait l'estimation d'environ  $t=500$  paramètres. La mise à jour d'une matrice de dimension  $(t \times t)$  qu'exigent à chaque itération les méthodes de minimisation numérique paraît alors impossible, surtout dans une optique où le choix du modèle final devrait s'effectuer de façon conversationnelle en essayant d'ajuster différents modèles.

- et aussi, parce que Jöreskog fait référence la plupart du temps à une décomposition "Spearmanienne" (du type (3.1)) de la matrice de variance. Ce type de décomposition pose un problème quand on cherche à calculer les "scores" (coordonnées des individus dans le système des axes factoriels), les composantes factorielles n'étant pas des combinaisons linéaires des variables initiales. Sachant que l'on ne peut faire intervenir des variables illustratives sur les plans factoriels, il y a forcément, comme le souligne Mulaik [Mulaik 1986], "ambiguïté dans le processus d'interprétation des indicateurs factoriels".

Ces raisons ne sont pas suffisantes, car, au-delà du problème posé par les calculs, la procédure proposée peut en réalité se passer de

l'hypothèse de normalité et opérer sur une décomposition du type de celles que l'on rencontre en Analyse en Composantes Principales<sup>1</sup>. La raison essentielle est qu'avec Jöreskog on respecte une philosophie anglo-saxonne éloignée de celle de l'analyse des données :

- pourquoi faire des hypothèses sur la matrice de variance associée à la population, alors que ces mêmes hypothèses peuvent être opérantes sur la matrice de variance de l'échantillon ?

- pourquoi faire de l'inférence, dès que le seul avantage qu'apporte le modèle probabiliste (normal) est une procédure de test à laquelle on ne croit que modérément, compte tenu des hypothèses complémentaires sur les distributions qui sont faites par ailleurs ? Remarquons que ces procédures sont d'ailleurs moins puissantes que celles plus empiriques basées sur des constats graphiques.

- pourquoi se préoccuper autant des individus alors que l'on ne se soucie guère des variables qui elles même doivent être considérées comme extraites d'un univers ?

Les remarques que nous faisons ici sur la démarche de Jöreskog rejoignent celles bien connues de L.Guttman sur l'approche du maximum de vraisemblance [Guttman 1985].

Dans les procédures que nous proposons en 3.3, on gardera certains éléments de l'approche de Jöreskog :

- avec la notion de la passerelle, on sera conduit à mélanger des variables connues (celles de la passerelle) et inconnues, "les premières étant choisies soigneusement de façon à extraire autant d'information que possible sur les secondes" [Jöreskog 1978, p.444].

- on s'efforcera de converger vers une solution finale en testant un ensemble d'hypothèses plus ou moins restrictives parmi lesquelles on opérera un tri en s'appuyant sur un indice d'ajustement. La solution

---

<sup>1</sup>. Dans COSAN, par exemple, où on peut imposer des contraintes aux éléments des inverses des matrices qui interviennent dans la spécification du modèle, la décomposition du type ACP de  $\Sigma$  ( $\Sigma=UDU'$  avec  $UU'=U'U=I$  et  $D$  diagonale) peut facilement être modélisée en considérant la représentation de Cayley d'une matrice orthogonale :  $U=(H+I)^{-1}(H-I)$ , où  $H$  est une matrice symétrique gauche (i.e.  $H'=-H$ ). On obtient alors des estimations des éléments de  $U$  (par l'intermédiaire de  $H$ ) et de  $D$  ainsi que les intervalles de confiance correspondants, en ajustant le modèle :

$$\Sigma = (H+I)^{-1}(H-I)D(H-I)'(H+I)^{-1}$$

finalement retenue ne découle pas seulement de considérations mathématiques ; elle dépendra aussi fortement, comme chez Jöreskog, des connaissances que l'on a sur le domaine étudié.

On pourra donc considérer que l'on adopte une attitude "confirmatoire", mais en ne donnant, et cela est bien dans l'esprit de l'analyse des données, qu'un petit rôle à l'inférence.

### 3.1.2. Analyse Conjointe de Plusieurs Tableaux de Données

L'objectif de l'analyse conjointe (par exemple la méthode STATIS) est la comparaison de  $m$  tableaux  $X_i$  concernant les mêmes individus et/ou les mêmes variables. Il s'agit ici d'un ensemble de techniques dont on trouvera une présentation rapide dans [Escoufier 1985] et une synthèse détaillée dans [Glaçon 1981] ; pour une interprétation probabiliste, on consultera [Foucart 1984].

De façon générale la communication s'effectue par l'intermédiaire d'un tableau résumé, synthèse des tableaux initiaux, qu'on appelle le *compromis*. Aux  $X_i (p_i \times n_i)$ ,  $i=1, \dots, m$ , étant associées les métriques  $M_i$  et  $D_{p_i}$  définies sur les espaces respectivement des individus et des variables, on doit choisir les éléments caractéristiques  $T(X_i)$  des tableaux  $X_i$  en fonction des éléments communs des triplets  $(X_i, M_i, D_{p_i})$ . On prendra, par exemple,  $T(X_i) = X_i' M_i X_i D_{p_i}$  ou  $T(X_i) = X_i D_{p_i} X_i' M_i$  selon que les tableaux font intervenir les mêmes individus ou les mêmes variables. On doit aussi choisir une mesure de la similarité entre tableaux, le plus souvent un produit scalaire entre éléments caractéristiques.

On peut ainsi, dans une première étape, bâtir une matrice symétrique  $C (m \times m)$  des similarités entre tableaux dont l'analyse, soit par des techniques de classification, soit par des techniques de positionnement multidimensionnel, permettra de reconnaître les tableaux proches au sens de la similarité choisie (analyse de l'interstructure). Dans une deuxième étape, on calcule le compromis comme une moyenne pondérée  $\sum a_i T(X_i)$  des éléments caractéristiques, le poids  $a_i$  accordé au  $i$ -ème tableau étant encore une question de choix (souvent les  $a_i$  sont pris égaux aux coordonnées -toutes positives- du premier vecteur propre de

C). La diagonalisation du compromis permet la représentation simultanée des éléments communs et non communs des différents tableaux (analyse de l'intrastructure) ; si la représentation des éléments non communs semble s'imposer de façon naturelle, celle des éléments communs pose un problème supplémentaire de choix.

Plus particulièrement, dans le cas où les différents tableaux contiennent des variables communes et non communes, on retiendra les sous-tableaux  $Z_i (p \times n_i)$ ,  $i=1, \dots, m$  contenant les variables communes seulement ; l'élément caractéristique du  $i$ -ème tableau est alors l'opérateur  $V_i M_i = Z_i D_{p_i} Z_i' M$  (le plus souvent  $M_1 = \dots = M_m = M$ ). L'analyse de la matrice des produits scalaires  $C$ , de terme général  $c_{ij} = \text{trace}(V_i M V_j M)$ , permettra l'étude de la proximité entre tableaux pris globalement. Si  $a_1, \dots, a_m$  sont les poids accordés aux différents tableaux (en présence d'une enquête de référence  $A$  on posera  $a_A = 1$  et  $a_j = 0$  ;  $j \neq A$ ), la diagonalisation du compromis ( $\sum a_i V_i M = U D_\lambda U'$  avec  $U U' = U' U = I$  et  $D_\lambda$  diagonale) permettra la représentation des individus des différents tableaux, qui n'est autre que celle fournie par l'A.C.P. du triplet  $(\tilde{Z}, M, D_p)$  où  $\tilde{Z} = [Z_1 | \dots | Z_m]$  et où  $D_p (\sum n_i \times \sum n_i)$  est la matrice diagonale obtenue en rangeant sur la diagonale les matrices  $a_i D_{p_i}$  ; les coordonnées des individus du  $i$ -ème tableau sur les axes factoriels sont données par les colonnes de  $U' M Z_i$ . Une représentation graphique possible des variables du  $i$ -ème tableau s'obtient en restreignant les composantes factorielles normées du triplet  $(\tilde{Z}, M, D_p)$  au sous-espace correspondant aux individus du  $i$ -ème tableau (ces coordonnées sont alors les lignes du tableau  $D_{1/\lambda} U' M V_i$ ). La représentation simultanée des variables de tous les tableaux s'obtient par simple superposition des représentations graphiques individuelles.

Il est connu que la méthodologie globale décrite s'avère particulièrement efficace quand les différents tableaux concernent les mêmes individus comme c'est le cas, par exemple, dans l'analyse des trajectoires ou dans l'analyse des correspondances "optimales" (voir section 4.5.). Si on se place dans une optique purement exploratoire, on peut y avoir recours dès que les tableaux concernent des variables ayant des significations indépendantes des individus considérés.

Or, dans le cadre des enquêtes psycho-sociales dont nous traitons la situation est bien différente : d'une part, on a des connaissances suffisamment précises du phénomène étudié avant d'entamer une étude

comparative et, d'autre part, la signification des variables dépend toujours de la population d'individus considérée. Nous n'aurons alors pas recours à l'Analyse Conjointe car, au delà de l'arbitraire inhérent à la représentation des variables et le peu d'intérêt du compromis (avec l'enquête de référence, il se réduit à un seul tableau  $Z_A$ ), la méthode semble peu adaptée à tenir compte du fait que des variables communes peuvent avoir des significations bien différentes. Il apparaît, ici en particulier, que le problème de l'analyse conjointe avec variables communes n'est pas simplement le problème "dual" (tel qu'il est présenté, par exemple, dans [Glaçon 1981]) de l'analyse avec individus communs.

Au niveau des calculs, l'Analyse Conjointe revient à positionner directement les individus des enquêtes particulières en tant que points supplémentaires sur le repère factoriel de l'enquête de référence. Les compensations, ayant lieu lors de l'application de coefficients inappropriés (coordonnées des vecteurs propres de la matrice d'inertie correspondant aux variables ayant des significations différentes) aux réponses individuelles, peuvent alors conduire à des résultats difficilement interprétables. Notons que l'utilisation du compromis n'est pas justifiée pour la raison supplémentaire que la matrice de variance conjointe contiendra des éléments (lignes et colonnes) qui sont des moyennes pondérées des covariances provenant de variables hétéroclites ; la diagonalisation de cette matrice risque d'être dépourvue d'intérêt concret.

Toutes ces critiques de l'analyse conjointe prise "au pied de la lettre", c'est à dire appliquée sans précautions, ne sont judicieuses que dans la mesure où, voulant dépasser la simple description des données dont on dispose, on cherche à bâtir des représentations respectant ce que l'on sait par ailleurs du phénomène. Le système social que l'on analyse à différentes époques est soumis à des perturbations qui peuvent provoquer, à certains moments, le désordre. Seules les structures dégagées aux périodes d'équilibre sont des caractéristiques "éclairantes" d'un système dynamique.

L'une des deux procédures que nous proposons pour traiter des données temporelles (paragraphe 3.2.4.) pourrait être assimilée à une analyse conjointe, où l'essentiel du travail ne résiderait plus dans le calcul, quasi-automatique, du compromis. Il résiderait dans tout un

ensemble d'aménagements exploitant au mieux la connaissance que l'on a du réel, connaissance où l'on voit se rejoindre, de façon paradoxale (cf remarques du début du paragraphe 3.2.5.), instabilité des variables et stabilité des structures.

### 3.1.3. L'analyse de redondance

L'objectif de cette technique<sup>1</sup>, qui ne traite pas directement de la comparaison entre structures factorielles, est l'analyse, au niveau du même tableau de données, d'un ensemble de variables par rapport à un autre ensemble pris comme référence.

Notons  $x^1, \dots, x^p$  un ensemble de variables mesurées sur  $n$  individus, partitionné en deux sous-ensembles :  $\{y^1, \dots, y^{p-q}\}$ ,  $\{z^1, \dots, z^q\}$ . Soit  $X(p \times n)$ ,  $Y((p-q) \times n)$  et  $Z(q \times n)$  les tableaux centrés correspondants :  $X' = [Y' | Z']$ . Soit encore,  $D_p$  étant la métrique diagonale des poids,  $V_{zz} = ZD_p Z'$ ,  $V_{yy} = YD_p Y'$  et  $V_{zy} = V'_{yz} = ZD_p Y'$ . L'analyse de redondance des variables  $y^j$  par rapport aux variables (de référence)  $z^i$  consiste à rechercher les combinaisons linéaires orthonormées (composantes redondantes)  $\xi = Z'a$  des variables  $z^i$  qui expliquent au mieux la variance des variables  $y^j$ , c'est à dire qui maximisent successivement, sous la contrainte d'orthogonalité, la quantité  $\xi' D_p Y' Y D_p \xi$ , somme de carrés des

---

<sup>1</sup> La technique doit son nom à Wollenberg [1977] qui s'est inspiré de l'indice de redondance de Stewart et Love (indice utilisé en Analyse Canonique, mesurant la variance moyenne des variables d'un ensemble expliquée par une composante canonique de l'autre ensemble). Une relation étroite existe entre la technique et l'"Analyse en Composantes Principales par rapport à des Variables Instrumentales" introduite par C.R. RAO [1964]. Nous en donnons ici quelques références :

- Rao, C. R. (1964). *The use and interpretation of principal components analysis in applied research*. Sankhya, the Indian Journal of Statistics, 26, 1, 329-358.
- van den Wollenberg, A. L. (1977). *Redundancy Analysis. An alternative to canonical correlation analysis*. Psychometrika, 42, 2, 207-220.
- Tyler, D. E. (1982). *On the optimality of the simultaneous redundancy transformations*. Psychometrika, 47, 1, 77-86.
- Israëls, A. Z. (1984). *Redundancy Analysis for qualitative variables*. Psychometrika, 49, 3, 331-346.

et encore : Robert et Escoufier [1976], Johansson [1981], Bonifas et al [1984], Lafosse [1985] et Israëls [1986], cités dans la bibliographie générale.

covariances entre  $\xi$  et  $y^1, \dots, y^q$ . Les facteurs  $a$  vérifient alors l'équation :

$$V_{ZZ}^{-1} V_{ZY} V_{YZ} a = \lambda a \quad (3.4)$$

On peut par la suite rechercher dans l'espace défini par les variables  $y^j$  les combinaisons linéaires  $\eta = Y'b$  les plus proches des composantes  $\xi$  et satisfaisant certaines contraintes. N'étant pas ici en analyse canonique, on ne peut exiger la bi-orthogonalité entre les variables  $\eta$  d'une part, et les variables  $\eta$  et  $\xi$  non homologues d'autre part. Aussi, n'exigera-t-on qu'une seule de ces conditions [Johansson 1981].

Notons  $\hat{Y}(q \times n)$  le tableau des  $D_p$ -projections des vecteurs  $y^j$  sur l'espace engendré par les lignes de la matrice  $Z$  ; on montre facilement que les composantes redondantes ne sont autres que les composantes principales normées du triplet  $(\hat{Y}, I, D_p)$ . L'analyse de redondance peut donc être considérée comme "l'analyse fine" (nous reprenons ici, dans un contexte légèrement différent, le terme utilisé par [Croquette et al. 1984]) de la variance expliquée :  $V_{YZ} V_{ZZ}^{-1} V_{ZY}$  ; mais on utilise dans l'espace des individus la métrique identité et non pas la métrique  $V_{YY}^{-1}$  comme en analyse canonique. On peut encore, comme le font Robert et Escoufier [Robert et Escoufier 1976], ramener l'analyse de redondance à la recherche d'une matrice carrée  $T(q \times q)$  telle que,  $M$  étant la métrique dans l'espace des individus, les ACP des triplets  $(Z, M, D_p)$  et  $(TY, M, D_p)$  conduisent à des résultats le plus identiques possibles.

On pourra juger de l'efficacité de la méthode en se reportant à l'exemple de la section 5.4. où on a effectué l'analyse de la redondance des variables relatives à la perception des risques par rapport à des variables-conflits. L'analyse de redondance a donc son utilité pour évaluer la puissance explicative d'un ensemble de variables de référence. Mais, dans le problème de communication qui nous intéresse ici, notons que l'on cherche à reconstruire dans l'espace de  $B$  des indicateurs équivalents à ceux de  $A$  plutôt qu'à analyser les variables de  $B$  par rapport à celles de  $A$  en général. On peut, bien-sûr, effectuer des rotations procrustéennes sur les composantes redondantes pour retrouver les indicateurs de référence, comme le propose Israëls [Israëls 1986] (de façon analogue à la technique que nous exposons en 3.2.5.), mais là encore on ne voit pas l'intérêt d'une rotation sur les composantes de la matrice  $\hat{Y}$  alors qu'on peut l'effectuer directement sur les composantes de la matrice  $Z$ .

### 3.2. Méthodologie proposée

On désire analyser les individus et les variables d'une enquête particulière B à l'aide de l'information contenue dans le repère factoriel de l'enquête de référence A ; il faut alors établir un repère "équivalent" dans l'espace factoriel significatif de B. Ici, l'équivalence est définie sur la base d'un nombre suffisant d'éléments communs devant être disposés de façon identique dans les deux repères. Cela concerne à la fois les coordonnées des variables quantitatives (actives ou non) sur les composantes factorielles et les coordonnées des modalités des variables qualitatives illustratives sur les axes factoriels correspondants.

Dans les procédures de communication entre enquêtes qui sont développées ici, nous proposons une méthodologie de construction de repères équivalents ; nous parlerons des précautions qu'il faut prendre, des hypothèses supplémentaires qu'il faut faire intervenir et plus généralement de ce qu'on a droit de faire lorsqu'on s'écarte de la situation "idéale" d'équivalence totale à laquelle on a fait référence précédemment. Par exemple, quand on obtient des repères ayant une structure similaire dans l'un seulement des deux espaces (celui des individus ou celui des variables), de quelle équivalence et de quelle passage d'information peut-on parler ?

On commence cette section en donnant des précisions sur la notion de passerelle : quels sont les principes que doit observer ce sous-ensemble de variables permettant la communication ? On décrit ensuite des techniques pour la construction d'une passerelle. On présente, enfin, deux procédures de communication : la première (méthode de la greffe) fait appel à la technique des points supplémentaires et à la régression linéaire ; la seconde s'appuie sur les techniques des rotations procrustéennes dans l'espace des variables.

#### 3.2.1. Notion de passerelle

Dans les deux procédures, la jonction entre les enquêtes A et B est établie à partir d'un sous-ensemble de variables communes ayant de

"bonnes" propriétés, qu'on appelle passerelle de A vers B. La passerelle ne fait pas l'objet d'une définition ; c'est une notion souple, issue des principes sur lesquels il faut s'appuyer pour la construire :

1) les variables (de la passerelle) doivent être peu nombreuses par rapport au nombre de variables initiales.

2) les variables doivent permettre de reconstruire au mieux l'information apportée par l'enquête de référence ; en particulier, si on s'intéresse aux indicateurs factoriels de référence, de reconstruire au mieux ces indicateurs.

3) les variables doivent avoir la même signification dans les deux enquêtes considérées. Ce principe implique que les coordonnées (covariances) des variables sur l'ensemble des composantes factorielles, réduit aux composantes significatives, puissent être considérées comme identiques.

4) les variables doivent être suffisamment différentes les unes des autres.

Comme il est montré plus loin, on est souvent amené à attacher moins d'importance à certains de ces principes et cela en fonction de la technique utilisée et de la nature du problème étudié.

Les deux premiers principes, qui sont en partie antagonistes, soulèvent le problème classique, en Analyse de Données, de sélection de variables (ici on est toutefois dans le cadre, peu habituel, de sélection en ACP) ; aussi, si les variables communes ne sont pas imposées par des critères propres à l'étude particulière menée, on a intérêt à utiliser, dans l'enquête de référence, une procédure optimale d'extraction de variables. Cette procédure, on le verra, a l'avantage de conduire à une passerelle qui respecte le quatrième principe.

Nous décrivons par la suite :

- des procédures optimales d'extraction de variables : ces variables, insérées dans les questionnaires, formeront la passerelle à condition d'observer le troisième principe.

- des procédures qui permettent, partant d'une première passerelle, d'éliminer les variables "indésirables" pour respecter, dans la mesure du possible, le troisième principe.

### 3.2.2. Sélection d'un sous-ensemble "optimal" de variables

Si en régression linéaire multiple et en analyse discriminante on dispose d'un certain nombre de techniques de sélection de variables, c'est, en partie, à cause de l'existence des critères "naturels" par rapport auxquels la sélection optimale est effectuée (le coefficient de corrélation multiple, l'erreur de prédiction moyenne...). En ACP ce que l'on peut exiger d'un sous-ensemble optimal de variables, est qu'il reproduise au mieux les caractéristiques générales de la totalité des données ; en particulier, qu'il rende possible la reconstruction des composantes principales.

Soit,

- $x^1, \dots, x^p$  l'ensemble de variables initiales centrées (tableau et métriques associées  $X(p \times n)$ ,  $M(p \times p)$ ,  $D_p(n \times n)$ ),
- $c^1, \dots, c^k$  le système orthogonal des  $k$  premières composantes principales jugées significatives issues de l'ACP du triplet  $(X, M, D_p)$  (tableau associé  $C(k \times n)$ ).

On cherche à extraire un sous-ensemble de variables  $(z^1, \dots, z^q / q > k)$ , (tableau associé  $Z(q \times n)$ ), le plus représentatif de  $(x^1, \dots, x^p)$ .

Notons :

- $y^1, \dots, y^{p-q}$ , l'ensemble des variables non sélectionnées (tableau  $Y((p-q) \times n)$ ) ; alors  $X' = [Y' | Z']$ ,
- $\|A\|^2 = \text{trace}(A^2)$ , pour toute matrice symétrique  $A$ ,
- $\rho_i(X, Y)$ , le  $i$ -ème coefficient de corrélation canonique entre les variables-lignes des tableaux  $X$  et  $Y$ ,
- $R^2(y ; x^1, \dots, x^j)$ , le carré du coefficient de corrélation multiple (c.c.m.) entre la variable  $y$  et les variables  $x^1, \dots, x^j$ .

Deux ensembles de propositions concernant l'extraction d'un sous-ensemble de variables ont été faits par Joliffe [Joliffe 1972, 1973, 1986] et McGabe [McGabe 1984]. Joliffe propose des procédures itératives descendantes où, partant de l'ensemble des variables, on élimine à chaque étape, après diagonalisation, les variables les plus corrélées avec les dernières composantes principales. McGabe introduit le concept de "variable principale", variable qui contient "autant d'information que possible" ; il cherche alors à obtenir un sous-ensemble de variables

(ses variables principales) ayant des propriétés analogues aux propriétés d'optimalité dont jouissent les composantes principales classiques. Pour cela, il passe en revue une douzaine de critères qui se ramènent tous à l'un des quatre critères suivants :

$$(C1) \max |v_{zz}| \text{ qui revient à } \min |v_{yy.z}|, \text{ car } |v_x| = |v_{zz}| |v_{yy.z}|.$$

$$(C2) \min \text{trace}(V_{yy.z}).$$

$$(C3) \min \|v_{yy.z}\|^2.$$

$$(C4) \max \sum_{i=1}^q \rho_i^2(Y,Z).$$

où,  $V$  pouvant être aussi une matrice de corrélation,  $v_{yy.z} = v_{yy} - v_{yz} v_{zz}^{-1} v_{zy}$  est la matrice de variance entre variables non sélectionnées sachant les variables sélectionnées.

Le premier critère exprime la maximisation de la variance généralisée des variables sélectionnées ; le deuxième (cas où  $V$  est une matrice de corrélation) revient à maximiser la quantité :  $\sum_{j=1}^p R^2(x^j ; z^1, \dots, z^q)$ , somme des carrés des coefficients de corrélation multiple de l'ensemble des variables par rapport aux variables sélectionnées ; avec le troisième, on minimise la variance non expliquée ( $\|A\|^2 = \sum a_{ij}^2$ ), et, avec le quatrième, on maximise une mesure de la proximité entre les espaces définis par les lignes des tableaux  $Y$  et  $Z$ .

Un critère voisin des précédents est proposé dans [Bonifas et al. 1984]<sup>1</sup> ; il s'agit d'une généralisation du critère proposé par [Robert et Escoufier 1976] dans le cas d'une métrique  $M$  quelconque. Sachant [Cailliez et Pagès 1976, ch. 14] que les ACP des triplets  $(X, M, D_p)$  et  $(Z, N, D_p)$  conduisent à des représentations identiques (homothétiques) des individus si et seulement si les opérateurs  $W_x = X' M X D_p$  et  $W_z = Z' N Z D_p$  sont égaux (proportionnels), les auteurs cherchent le sous-ensemble de variables  $z^1, \dots, z^q$  et la "métrique"  $N$  qui maximisent le cosinus entre les opérateurs  $W_x$  et  $W_z$  :

$$(C5) \max RV(W_x, W_z) = \frac{\text{Trace}(W_x W_z)}{\|W_x\| \|W_z\|}$$

<sup>1</sup> on trouve un critère analogue dans le cas d'une Analyse Factorielle des Correspondances dans [Dambroise et al. 1987].

Lorsque  $M=I$ , on sait [Robert et Escoufier 1976 p.261] que, pour  $z^1, \dots, z^q$  fixés, la maximisation de  $RV(W_X, W_Z)$  par rapport à  $N$  conduit à la solution offerte par l'analyse de redondance (cf. section 3.1.3.) des variables  $x^i$  par rapport aux variables  $z^j$ . On a au maximum :

$$RV(W_X, W_Z) = \frac{\|V_{XZ}V_{ZZ}^{-1}V_{ZX}\|^2}{\|V_{XX}\|^2}.$$

Ici, on cherche donc à maximiser la norme de la variance expliquée (à rapprocher au critère C3).

Avec les critères cités, qui ne font pas intervenir directement les  $k$  composantes significatives mais seulement des matrices de variance entre variables initiales, il y a de fortes chances qu'on retienne des variables peu explicatives de ces composantes. Le critère C1, par exemple, conduit à sélectionner des variables  $z^j$  bien diversifiées (cas où  $V$  est une matrice de corrélation) : moins les variables  $z^j$  sont corrélées, plus  $|R_{ZZ}|$  est grand, le maximum étant atteint pour  $R_{ZZ}=I$  ; c'est en ce sens seulement que les  $z^j$  sont représentatifs de la totalité.

Un critère, bien différent, qui tient compte plus directement de la représentation des individus sur les  $k$  composantes significatives est proposé par Krzanowski [Krzanowski 1987]. Soit  $D(k \times n)$  le tableau des  $k$  premières composantes principales issues de l'ACP du triplet  $(Z, M_Z, D_p)$ , où  $M_Z$  est la restriction de  $M$  sur les  $z^i$ . Pour mesurer l'écart entre les triplets  $(X, M, D_p)$  et  $(Z, M_Z, D_p)$ , relativement à leur sous-espaces principaux de dimension  $k$ , Krzanowski utilise l'écart de procruste<sup>1</sup> entre les

---

<sup>1</sup> Rappel : Soit  $E_1$  et  $E_2$  deux nuages de points appariés, munis des poids  $p_i$  et soit  $P$  une isométrie (translation plus rotation) qui s'applique à  $E_2$ . Le minimum par rapport à  $P$  de la somme pondérée des carrés des distances entre points homologues de  $E_1$  et de  $P(E_2)$  est appelé écart de procruste entre  $E_1$  et  $E_2$ .

Matriciellement, si  $(X_1, M_1, D_p)$  et  $(X_2, M_2, D_p)$  sont les triplets correspondants (pour simplifier, les tableaux  $X_1(p_1 \times n)$  et  $X_2(p_2 \times n)$  sont supposés centrés ;  $p_2 < p_1$ ) et si  $P(p_1 \times p_2)$  est une matrice de rotation, c'est à dire telle que  $P'M_1P=M_2$ , l'écart de procruste  $\Delta^2(E_1, E_2)$  s'écrit :

$$\Delta^2(E_1, E_2) = \min_P \{ \text{trace} [(X_1 - PX_2)'M_1(X_1 - PX_2)D_p] \}$$

nuages de points construits à partir des tableaux des composantes principales  $C(k \times n)$  de  $X$  et  $D(k \times n)$  de  $Z$  :

$$\Delta^2 = \min_P \{ \text{trace} [(C-PD)'(C-PD)D_p] \} = \sum_{i=1}^k [\lambda_i^X + \lambda_i^Z + 2\lambda_i^{CD}]$$

où,

$P'P=I$  et  $\lambda_i^X, \lambda_i^Z, \lambda_i^{CD}$  sont les  $i$ -èmes valeurs singulières des matrices respectivement  $V_{XX}, V_{ZZ}$  et  $CD'$ .

Avec le critère de Krzanowski on cherche donc le sous-ensemble de variables  $z^1, \dots, z^q$  minimisant le critère :

$$(C6) \min \sum_{i=1}^k [\lambda_i^X + \lambda_i^Z + \lambda_i^{CD}].$$

Selon l'auteur, ce critère s'avère nettement supérieur aux précédents dans le cas où la représentation des individus sur les plans principaux révèle l'existence de groupes.

Pour tous les critères précédents, l'exploration des  $p!/(q!(p-q)!)$  sous-ensembles paraît impossible pour  $p$  et  $q$  élevés. On est ainsi amené à considérer des procédures de sélection pas à pas ascendantes ou descendantes. On bâtit facilement de telles procédures en exploitant les formules classiques de passage :

- de l'inverse d'une matrice symétrique à l'inverse de la matrice obtenue en ajoutant ou en enlevant une ligne et une colonne (voir la formule 3.9 plus loin),
- de la décomposition en valeurs singulières (d.v.s.) d'une matrice symétrique à la d.v.s. de la matrice obtenue en enlevant une ligne et une colonne (voir [Krzanowski 1987]).

Les indicateurs de référence, que l'on veut reconstruire ici à partir d'un sous-ensemble de variables, ne sont pas nécessairement des composantes principales (ils se situent, néanmoins, dans un sous-espace principal significatif). Pour les faire intervenir directement dans la sélection des variables, sans pour autant restreindre leur reconstruction dans le sous-espace principal de dimension  $k$  du tableau  $Z$ , nous proposons ici deux procédures inspirées de la régression pas à pas.

■ Dans la première, on compare à chaque itération les coefficients de corrélation multiple partiels entre les variables sélectionnées et les composantes de référence sachant les variables sélectionnées :

étape 0 : Poser  $c_0^\ell = c^\ell$  ;  $\ell=1, \dots, k$ .

———— itérer pour  $m=1$  à  $q$  sur les deux étapes suivantes ————

étape 1 : chercher  $\max_j R^2(x^j ; c_{m-1}^1, \dots, c_{m-1}^k)$ , obtenu pour  $j=j^*$ .  
 $x^{j^*}$  est alors sélectionné : Poser  $z^m = x^{j^*}$ .

étape 2 : éliminer des  $c_{m-1}^\ell$  leurs parties prévisibles par régression linéaire sur  $z^m$ , en effectuant les  $k$  régressions correspondantes :  $c_m^\ell = c_{m-1}^\ell - b_m^\ell z^m$  ;  $\ell=1, \dots, k$ .

■ Dans la deuxième procédure, ce sont les cosinus entre les opérateurs de projection  $W_z$  et  $W_c$  associés aux espaces engendrés respectivement par les variables initiales et les composantes factorielles que l'on compare à chaque itération. On a :  $W_z = Z'V_{ZZ}^{-1}ZD_p$  et  $W_c = C'V_{CC}^{-1}CD_p$ . On cherche alors à maximiser :

$$RV(W_c, W_z) = \frac{\text{trace}\{V_{CC}^{-1}V_{CZ}V_{ZZ}^{-1}V_{ZC}\}}{(kq)^{\frac{1}{2}}} = \frac{\sum_{\ell=1}^k \rho_\ell^2(Z, C)}{(kq)^{\frac{1}{2}}}$$

qui, dans la mesure où les composantes sont orthogonales, s'écrit encore :

$$RV(W_c, W_z) = \frac{\text{trace}\{R_{CZ}R_{ZZ}^{-1}R_{ZC}\}}{(kq)^{\frac{1}{2}}} = \frac{\sum_{\ell=1}^k R^2(c^\ell ; z^1, \dots, z^q)}{(kq)^{\frac{1}{2}}}$$

Pour la mise en place de la procédure ascendante correspondante, on réécrit  $\text{trace}\{R_{CZ}R_{ZZ}^{-1}R_{ZC}\} = \text{trace}\{R_{ZZ}^{-1}R_{ZC}R_{CZ}\}$  ; on calcule une fois pour toutes la matrice  $R_{CZ}R_{CZ}$  et, à chaque étape, on utilise la formule de passage (3.9).

Les deux procédures ont été programmées avec possibilité d'imposer ou d'exclure certaines variables. Dans l'exemple de la figure 3.1, on constate qu'on arrive à approcher de façon satisfaisante (au sens du  $R^2$ ) les 4 premières composantes principales d'une A.C.P. portant sur 100 variables avec le cinquième seulement des variables initiales.

Soit  $\hat{C} = V_{CZ}V_{ZZ}^{-1}C$  le tableau des approximations  $\hat{c}^\ell$  obtenues par la régression linéaire des composantes  $c^\ell$  par rapport à des variables  $z^j$  sélectionnées comme précédemment. Nous verrons plus loin, en discutant sur la méthode de la "greffe", que la perte d'information qui accompagne

METHODE PAR COMPARAISON DES  
COEFFICIENTS DE CORRELATION  
MULTIPLE PARTIELS

1. RÉTABLIR LA PEINE DE MORT
2. EMBAUCHER DANS LES SERVICES PUBLICS
3. CENSURER CERTAINS LIVRES
4. POUR LA FAMILLE
5. MONTEE DE LA VIOLENCE  
c.c.m. 0.819, 0.654, 0.461, 0.494
6. POUR LA PATRIE
7. MOINS DE ROBOTS
8. POUR LES 35 HEURES
9. SOUTENIR LES PACIFISTES
10. MAINTENIR LA CROISSANCE  
c.c.m. 0.892, 0.805, 0.662, 0.703
11. CONFIANCE EN LA JUSTICE
12. VOTE MUNICIPAL AUX ETRANGERS
13. ETUDIANTS PARASITES
14. GRANDES FORTUNES RESPECTABLES
15. UTILISER L'ENERGIE SOLAIRE  
c.c.m. 0.918, 0.852, 0.735, 0.812
16. CONTRE LA COHABITATION
17. TROP DE FAUX CHOMEURS
18. MEDECINE GRATUITE
19. ACHETER FRANCAIS
20. LA PROPRIETE C'EST LE VOL  
c.c.m. 0.942, 0.898, 0.780, 0.844

METHODE PAR COMPARAISON  
DES COSINUS ENTRE  
OPERATEURS DE PROJECTION

1. RÉTABLIR LA PEINE DE MORT (1)
2. EMBAUCHER DANS LES SERVICES PUBLICS (2)
3. POUR LA FAMILLE (4)
4. POUR LA PATRIE (6)
5. MOINS DE ROBOTS (7)  
c.c.m. 0.786, 0.678, 0.588, 0.448
6. UTILISER L'ENERGIE SOLAIRE (15)
7. CONFIANCE EN LA JUSTICE (11)
8. CENSURER CERTAINS LIVRES (3)
9. SOUTENIR LES PACIFISTES (9)
10. MAINTENIR LA CROISSANCE (10)  
c.c.m. 0.837, 0.760, 0.709, 0.753
11. CONTRE LA COHABITATION (16)
12. MEDECINE GRATUITE (18)
13. MONTEE DE LA VIOLENCE (5)
14. ETUDIANTS PARASITES (13)
15. ENTREPRISES NATIONALISEES AUSSI EFFICACES  
c.c.m. 0.900, 0.847, 0.789, 0.813
16. HOMMES POLITIQUES INTEGRES
17. LA PROPRIETE C'EST LE VOL (20)
18. POLLUTION PREOCCUPANTE
19. GRANDES FORTUNES RESPECTABLES (14)
20. POUR LES 35 HEURES (8)  
c.c.m. 0.915, 0.900, 0.825, 0.864

Enquête "Agoramétrie 1985" :

- nombre de variables dans l'ACP :  $p=100$
- nombre de composantes à reconstruire :  $k=4$
- nombre de variables de la passerelle :  $q=20$
- nombre d'individus  $n=1028$ .

c.c.m. : coefficient de corrélation multiple entre chacune des 4 composantes et les variables déjà citées.

A noter que, sur les 20 variables sélectionnées, 17 sont les mêmes pour les deux techniques.

FIGURE 3.1 : Extraction des variables optimales

le passage de  $C$  à  $\hat{C}$  est relativement petite. En effet, pour des c.c.m. de l'ordre de 0.9 et pour le type et la taille des données traitées, les coordonnées, aussi bien des variables sur les composantes factorielles que celles des modalités significatives des variables qualitatives illustratives sur les axes factoriels correspondants, sont sensiblement identiques sur les indicateurs factoriels initiaux et reconstruits.

#### Remarque

Des variables bien corrélées aux variables déjà introduites, ou encore des variables expliquant les mêmes parties des composantes factorielles de référence, ne sont pas sélectionnées dans la passerelle quand on a recours à ces procédures de sélection pas à pas ; c'est pourquoi ces procédures permettent de respecter le quatrième principe de "diversité", principe qu'il faut observer surtout dans la procédure de communication par rotation.

En régression, si deux variables contribuent de la même façon à la variable à expliquer (ici la composante factorielle  $c^l$ ), elles ne créent pas de problème dans le calcul des coefficients de régression (ni par conséquent dans le calcul des reconstructions  $\hat{c}_i^l$  des coordonnées des individus) dans la mesure où elles sont peu corrélées entre elles<sup>1</sup> (l'introduction d'une variable  $z^j$ , redondante par rapport à la variable  $z^i$  dans le modèle, donnera, en général, un nouveau coefficient  $\hat{b}_j^l$  non nul, dont l'effet dans la formule de reconstruction sera compensé par une baisse en valeur absolue du coefficient  $\hat{b}_i^l$ ). Par contre, dans les rotations procrustéennes, ces variables peuvent nuire à la qualité des résultats : en donnant un poids trop grand à une catégorie de variables, on introduit des distorsions dans la répartition des observations sur l'axe correspondant à la composante reconstruite.

#### 3.2.3. Elimination de variables

Dans les deux procédures de communication, par régression ou par rotation, la reconstruction du repère factoriel dans  $B$  s'effectuera sur la base des variables de la passerelle : soit en construisant les compo-

---

<sup>1</sup>ce qui est le cas dans les enquêtes psycho-sociales traitées ici.

santes factorielles comme des combinaisons linéaires des variables de la passerelle (régression), soit en effectuant une rotation des composantes factorielles ayant comme cible une matrice constituée uniquement des coordonnées de ces mêmes variables sur les composantes de référence.

On a vu que souvent des variables communes ne sont pas équivalentes (perçues de la même façon), quoiqu'ayant le même libellé dans A et dans B. Dans toute procédure de communication ces variables ne devront pas participer à la reconstruction des composantes factorielles. Pour chaque composante, il faut alors exclure de la passerelle les variables dont les parties expliquées par la composante ne sont pas identiques dans les deux enquêtes. Pour ce faire, on peut construire les composantes à partir de toutes les variables communes, et éliminer par étape les variables qui semblent avoir des parties expliquées différentes. Voulant éviter cette procédure itérative qui pose problème dès que plusieurs variables sont inappropriées, nous nous sommes contenté d'une procédure moins économe en travaillant directement sur les matrices de variance ; il y a alors un faible risque d'exclure des variables qui, ayant changé de signification, pourraient être néanmoins conservées pour reconstruire certaines des composantes.

Opérer de cette façon est justifié par le fait que les parties des variables expliquées par les composantes factorielles de référence dépendent des liaisons que les variables entretiennent les unes avec les autres. Si les matrices de variance entre les variables communes sont sensiblement différentes, il est fort probable que ces différences se retrouvent dans les parties expliquées de certaines variables. Ici, on exploite le fait que la principale source de covariation entre les variables est leur partie expliquée par les composantes significatives, appelée "partie commune" par les factoralistes classiques.

Pour aboutir à la passerelle finale on s'appuiera donc sur les équivalences suivantes entre les matrices de variance  $V_A$  et  $V_B$  associées aux variables communes dans les enquêtes A et B respectivement :

- Egalité des matrices de variance :

$$V_A = V_B ; \quad (3.5)$$

cette égalité exprime une identité structurelle totale.

- Egalité des matrices de corrélation :

$$R_A = R_B ; \quad (3.6)$$

seule l'intensité des liaisons entre variables est ici considérée comme importante : on s'affranchit des homothéties au niveau des différences individuelles.

- Similitude des matrices de corrélation :

$$R_A - \text{diag}(R_A) = \alpha(R_B - \text{diag}(R_B)) , \alpha > 0 ; \quad (3.7)$$

on ne s'intéresse ici qu'à la forme des liaisons entre variables.

Notons qu'avec ces trois équivalences les matrices  $R_A$  et  $R_B$  ont les mêmes vecteurs propres. D'autres équivalences pourraient être retenues, mais seules les précédentes sont pour l'instant opérationnelles.

Ainsi, partant d'un ensemble de  $p$  variables communes, on écartera de la passerelle, par une procédure descendante, à chaque étape, la variable qui conduit, après élimination, à la plus petite valeur de la "distance" entre les matrices de dispersion.

Pour les trois équivalences, on bâtit facilement des procédures d'élimination basées sur la distance des moindres carrés :  $d = \|V_A - V_B\|^2$ . Une telle procédure a été retenue pour l'équivalence (3.7). Pour les équivalences (3.5) et (3.6), on propose deux procédures alternatives qui trouvent une justification statistique dans le cas où les populations associées à **A** et **B** sont indépendantes et multinormales.

Equivalence (3.5) : Soit  $n_A$  et  $n_B$  les tailles des échantillons de **A** et **B**,  $\Sigma_A$  et  $\Sigma_B$  étant les matrices de variance théoriques associées aux  $q$  variables communes entre **A** et **B**. Pour tester l'hypothèse  $H_0 : \Sigma_A = \Sigma_B$ , on utilise la statistique de Box (cf. [Krishnaiah 1980, p.526-528]).

$$M = C_t \log \frac{|V_A|^{n_A} |V_B|^{n_B}}{|V|^{n_A+n_B}} \quad (3.8)$$

où :

- $V$  est la matrice de variance conjointe :  $V = 1/(n_A+n_B)(n_A V_A + n_B V_B)$
- $C_t$  une constante qui dépend du nombre d'observations (on peut poser  $C_t=1$ )

Sous  $H_0$ ,  $M$  est distribuée comme un khi-2 à  $q(q+1)/2$  degrés de

liberté<sup>1</sup>. Pour  $q$  élevé, Box suggère la transformation de  $M$  en un  $F$  de Fisher.

En utilisant  $M$  comme indice de distance, on établit une procédure d'élimination de variables : on calcule les inverses initiaux des  $(q \times q)$  matrices  $V_A, V_B$  et  $V$  par une procédure de triangularisation qui fournit comme résultat intermédiaire les déterminants de ces mêmes matrices. Par la suite, on itère en s'appuyant sur les deux résultats classiques suivants:

$$- |V_j| = v_{jj} |V|$$

où :  $V_j$  est la matrice obtenue en supprimant les  $j$ -ème ligne et colonne de  $V$  et  $v_{jj}$  le  $j$ -ème terme diagonal de  $V^{-1}$ .

$$- \text{Si } V = \begin{bmatrix} U & u \\ u' & d \end{bmatrix} \text{ et } V^{-1} = \begin{bmatrix} W & w \\ w' & e \end{bmatrix} \text{ où, } u \text{ et } w \text{ sont des vecteurs,}$$

$$\text{alors : } U^{-1} = W \left[ I + \frac{uw'}{1-u'w} \right] \quad (3.9)$$

Les itérations s'arrêtent quand la diminution de la valeur de  $M$  entre deux étapes successives est petite par rapport à la diminution du nombre de degrés de liberté. Un inconvénient de la méthode est l'inadéquation de  $M$  quand on s'écarte de la multinormalité.

Equivalence (3.6). Soit  $\rho_A$  et  $\rho_B$  les coefficients de corrélation théoriques entre deux variables dans  $A$  et  $B$ . Pour tester l'hypothèse  $H_0 : \rho_A = \rho_B$ , on utilise la statistique [Kshirsagar 1972, p.91-92] :

$$U = \frac{\tanh^{-1} r_A - \tanh^{-1} r_B}{\sqrt{\frac{1}{n_A - 3} + \frac{1}{n_B - 3}}} \quad (3.10)$$

---

<sup>1</sup>Comme il est montré dans [Manly et al. 1987], la statistique  $M$  qui n'est autre que celle du rapport de vraisemblance, peut être décomposée en :  $M = M_1 + M_2 + M_3$ , où les statistiques  $M_3, M_2, M_1$  servent à tester successivement les trois hypothèses emboîtées : matrices de corrélation égales, matrices de variance proportionnelles, matrices de variance égales.

où  $r_A$  et  $r_B$  sont les coefficients de corrélation empiriques. Sous  $H_0$ ,  $U$  suit une loi normale centrée réduite.

Soit  $U_{ij}$  la valeur de  $U$  pour les variables  $x^i$  et  $x^j$ , et :

$$u_i = \sum_{j=1}^q U_{ij}^2 \quad (3.11)$$

Tous les  $u_i$ ,  $i=1, \dots, q$ , suivent une loi de  $\chi^2$  ayant le même nombre de degrés de liberté ; on établit une procédure d'élimination en écartant à chaque étape la variable associée au plus grand des  $u_i$  ; les itérations s'arrêtent quand tous les  $U_{ij}$  sont non significatifs.

Les  $u_i$  étant décomposables en une somme des  $U_{ij}$ , on peut accompagner les itérations de la procédure par des graphiques illustratifs des écarts entre couples de variables (voir figure 3.2) ; on pourra ainsi décider à l'oeil de l'arrêt des itérations. Par ailleurs pour de petites et moyennes valeurs de  $r$ , comme  $\tanh^{-1} r \approx r$ , l'indice (3.11) revient à un indice de moindres carrés directement sur les corrélations.

Equivalence (3.7). On ne dispose pas ici d'indice statistique. A chaque itération, on calcule  $\alpha$  qui minimise la quantité :

$$\text{trace}[(R_A - \text{diag}(R_A) - \alpha(R_B - \text{diag}(R_B)))]^2.$$

On a :

$$\alpha = \frac{\text{trace}[(R_A - \text{diag}(R_A))(R_B - \text{diag}(R_B))]}{\text{trace}[R_B - \text{diag}(R_B)]^2} \quad (3.12)$$

On élimine la variable  $x^i$  pour laquelle  $\sum_{\substack{j=1 \\ j \neq i}}^q (r_A^{ij} - \alpha r_B^{ij})^2$  est maximum.

Dans la pratique on utilise le plus souvent le critère (3.6). Sur l'exemple qui apparaît dans la figure (3.2), le critère (3.5) a donné les mêmes résultats que le critère (3.6) pour les deux premières variables seulement qui sont manifestement les seules à "sortir du lot".

**Remarque** : il s'agit dans ces procédures de n'éliminer que les variables qui, à l'évidence, se comportent différemment dans les deux populations : il faut réaliser un compromis entre nombre minimum de variables dans la passerelle et stabilité des significations (proximité entre  $V_A$  et  $V_B$ ).

41 VARIABLES

	L	C	D	E	C	F	E	T	L	O	R	I	D	P	A	P	P	E	P	C	M	R	C	G	R	C	S	N	C	A	E	A	P	M	S	E	O	A	G	F	H	T	H													
MOYENNE 84																																																								
MOINS																																																								
MOYENNE 81																																																								
- 26																																																								
- 12																																																								
0 16																																																								
- 60																																																								
0 14																																																								
0 06																																																								
- 28																																																								
0 02																																																								
- 12																																																								
- 11																																																								
- 09																																																								
- 21																																																								
- 06																																																								
- 13																																																								
0 12																																																								
- 10																																																								
- 20																																																								
- 05																																																								
0 07																																																								
0 03																																																								
0 32																																																								
0 01																																																								
0 05																																																								
0 04																																																								
0 10																																																								
0 03																																																								
- 40																																																								
0 28																																																								
- 27																																																								
0 09																																																								
- 36																																																								
- 11																																																								
- 43																																																								
- 24																																																								
- 33																																																								
- 04																																																								
- 21																																																								
- 19																																																								
- 19																																																								
0 24																																																								
- 01																																																								

Enquêtes : Agoramétrie Mars 1981 (n=1543, p=76),  
 Agoramétrie Juin 1984 (n=1021, p=85).

nombre de variables communes q=41.

La valeur absolue de la différence entre les éléments homologues des deux matrices de corrélation a été codée selon la table suivante :

$d =  r_{ij}^{M81} - r_{ij}^{J84} $	symbole
$d \leq 0.05$	:
$0.05 < d \leq 0.10$	I
$0.10 < d \leq 0.15$	X
$0.15 < d \leq 0.20$	■
$0.20 < d$	■

A chaque étape la variable sortante (réinsertion ultérieure possible) est biffée et marquée par une étoile

FIGURE 3.2 : Elimination pas à pas des variables ayant des significations différentes (Suite...)

40 VARIABLES

MOYENNE 84  
MOINS  
MOYENNE 81

L	C	D	E	C	F	E	T	L	C	P	A	P	P	E	C	C	M	R	C	R	C	S	N	C	A	E	A	P	M	S	E	O	A	G	F	H	T	H				
I	D	I	G	O	O	T	R	I	O	R	I	O	O	N	O	O	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	
B	E	U	N	D	L	O	C	M	N	D	N	S	N	I	G	E	N	N	H	E	A	M	O	Y	C	R	O	R	E	A	M	O										
R	E	S	U	L	F	C	D	V	I	T	S	L	R	O	R	T	T	N	S	P	M	T	P	S	R	T	D	L	E	R	U	O	P									
N	E	S	I	E	I	A	T	R	P	U	R	R	T	D	U	A	I	A	E	P	G	T	R	I	R	L	D	C	E	E	E	E	E	E	E	E	E	E	E	E		
A	U	X	E	E	D	N	L	R	O	Y	I	A	A	C	P	E	O	E	E	D	I	E	A	E	E	P	N	R	P	R	C	E	O									
V	C	I	R	M	E	T	P	U	S	O	R	R	N	N	H	X	S	V	S	I	L	A	M	O	L																	
R	E	T	R	J	F	P	M	E	R	N	S	A	A	A	R	A	I	L	E	N	S	E	E	O	J	I	D	S	R	N	E	M	O	T								
T	A	E	U	R	A	M	R	N	A	O	P	T	T	M	A	R	S	L	I	N	G	I	N	S	L	R	S	O	I	L	A	D	E	N								
E	I	V	S	A	R	I	O	B	U	R	R	A	I	V	R	S	I	V	A	O	N	S	E	S	A	L	S	R	I	N	F	I	A	T	N							
M	R	E	T	P	A	G	T	G	R	O	E	I	L	L	A	I	A	T	R	N	S	E	R	E	I	R	E	E	E	E	E	E	E	E	E	E	E	E	E	E		
N	S	U	C	E	I	E	G	A	T	V	C	E	L	E	G	E	R	S	U	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E		
T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T	S	E	T

- 26 LIBERALISER L'AVORTEMENT
- 12 CONSTRUIRE DES CENTRALES NUCLEAIRES
- O 16 DIEU EXISTE
- 60 EGALISER LES REVENUS
- O 14 \*EBNFIANEE-EN-LA-JUSTICE-----
- O 06 POUR LA FORCE DE FRAPPE
- 28 ETUDIANTS PARASITES
- O 02 TROP DE TRAVAILLEURS IMMIGRES
- 12 LIMITER LES HERITAGES
- 11 CONTRE LA PORNOGRAPHIE
- 09 PRIS POUR ABRUTIS A LA TELEVISION
- 21 AIDER LES PAYS SOUS DEVELOPPES
- 06 POLLUTION PREOCCUPANTE
- 13 POUR LA PATRIE
- O 12 ENCOURAGER LA NATALITE
- 10 POUR LA FAMILLE
- 20 CONTRE LE TRAVAIL
- 05 CONTRE LE MARIAGE
- O 07 MAINTENIR LA CROISSANCE
- O 03 REDUIRE LES DEPENSES MILITAIRES
- O 32 CENSURER CERTAINS LIVRES
- O 05 RESPECTER LES CONVENANCES
- O 04 CHOMAGE ANGOISSANT
- O 10 SENTIMENT D'INSECURITE
- O 03 NE PAS HESITER A S'ENDETTER
- 40 CRISE DE L'ENERGIE PREOCCUPANTE
- O 28 AUGMENTER LE PRIX DE L'ESSENCE
- 27 UTILISER L'ENERGIE SOLAIRE
- O 09 ACCIDENTS DU TRAVAIL RARES
- 36 POUR LES 35 HEURES
- 11 MONTEE DE LA VIOLENCE
- 43 SYNDICATS INDISPENSABLES
- 24 ON N'APPREND PLUS RIEN A L'ECOLE
- 33 ORDINATEURS MENACENT NOS LIBERTES
- 04 ACHETER FRANCAIS
- 21 GUERRE MONDIALE INEVITABLE
- 19 FRAUDER LE FISC
- 19 HOMOSEXUELS COMME LES AUTRES
- O 24 TROP DE FONCTIONNAIRES
- 01 HOMMES POLITIQUES INTEGRES

REINSERTIONS : SENTIMENT D'INSECURITE

A LA PLACE DE POUR LES 35 HEURES

25 VARIABLES

MOYENNE 84  
MOINS  
MOYENNE 81

L	D	E	C	A	P	P	E	C	C	R	C	R	S	N	C	A	E	M	S	A	G	F	H	T	H																
I	T	O	I	O	N	O	O	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E																
B	E	U	N	D	L	O	C	M	N	D	N	S	N	I	G	E	N	N	H	E	A	M	O	Y	C	R	O	R	E	A	M	O									
R	E	S	U	L	F	C	D	V	I	T	S	L	R	O	R	T	T	N	S	P	M	T	P	S	R	T	D	L	E	R	U	O	P								
N	E	S	I	E	I	A	T	R	P	U	R	R	T	D	U	A	I	A	E	P	G	T	R	I	R	L	D	C	E	E	E	E	E	E	E	E	E	E	E		
A	U	X	E	E	D	N	L	R	O	Y	I	A	A	C	P	E	O	E	E	D	I	E	A	E	E	P	N	R	P	R	C	E	O								
V	C	I	R	M	E	T	P	U	S	O	R	R	N	N	H	X	S	V	S	I	L	A	M	O	L																
R	E	T	R	J	F	P	M	E	R	N	S	A	A	A	R	A	I	L	E	N	S	E	E	O	J	I	D	S	R	N	E	M	O	T							
T	A	E	U	R	A	M	R	N	A	O	P	T	T	M	A	R	S	L	I	N	G	I	N	S	L	R	S	O	I	L	A	D	E	N							
E	I	V	S	A	R	I	O	B	U	R	R	A	I	V	R	S	I	V	A	O	N	S	E	S	A	L	S	R	I	N	F	I	A	T	N						
M	R	E	T	P	A	G	T	G	R	O	E	I	L	L	A	I	A	T	R	N	S	E	R	E	I	R	E	E	E	E	E	E	E	E	E	E	E	E	E	E	
N	S	U	C	E	I	E	G	A	T	V	C	E	L	E	G	E	R	S	U	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	
T	T	C	E	E	R	S	U	D	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	

- 26 LIBERALISER L'AVORTEMENT
- O 16 DIEU EXISTE
- 28 ETUDIANTS PARASITES
- 11 CONTRE LA PORNOGRAPHIE
- 21 AIDER LES PAYS SOUS DEVELOPPES
- 06 POLLUTION PREOCCUPANTE
- 13 POUR LA PATRIE
- O 12 ENCOURAGER LA NATALITE
- 20 CONTRE LE TRAVAIL
- 05 CONTRE LE MARIAGE
- O 03 REDUIRE LES DEPENSES MILITAIRES
- O 32 CENSURER CERTAINS LIVRES
- O 05 \*RESPECTER LES CONVENANCES-----
- O 10 SENTIMENT D'INSECURITE
- O 03 NE PAS HESITER A S'ENDETTER
- 40 CRISE DE L'ENERGIE PREOCCUPANTE
- O 28 AUGMENTER LE PRIX DE L'ESSENCE
- 27 UTILISER L'ENERGIE SOLAIRE
- 11 MONTEE DE LA VIOLENCE
- 43 SYNDICATS INDISPENSABLES
- 04 ACHETER FRANCAIS
- 21 GUERRE MONDIALE INEVITABLE
- 19 FRAUDER LE FISC
- 19 HOMOSEXUELS COMME LES AUTRES
- O 24 TROP DE FONCTIONNAIRES

FIGURE 3.2 : (...Suite)

### 3.2.4. Technique par régression (méthode de la "greffe")

La façon la plus directe d'analyser les individus de **B** sur la base du repère factoriel de référence de **A** consiste, tout simplement, à les projeter en tant que points supplémentaires sur les axes factoriels de ce repère.

Ne disposant pas dans **B** de la totalité des variables de **A**, la méthode par régression propose de remplacer les composantes factorielles initiales de **A** par leurs projections (approximations par régression) sur l'espace engendré par les variables de la passerelle. Pour effectuer la "greffe", on applique alors la "formule de reconstruction" :

formule de reconstruction : la coordonnée d'un individu  $i$  (que ce soit un individu de **A** ou de **B**) sur le  $\ell$ -ème indicateur factoriel  $c^\ell$  est approchée par la formule [Nhun Fat 1984] :

$$\hat{c}_i^\ell = \sum_{j=1}^q b_j^\ell \frac{z_i^j - m_j}{s_j} \quad (3.13)$$

où,

- $b_j^\ell$  est le coefficient de régression à appliquer dans **A** à la  $j$ -ème variable de la passerelle, après centrage et réduction, pour reconstruire la composante  $c^\ell$ ,  $j=1, \dots, q$  ;  $\ell=1, \dots, k$ ,
- $m_j$  et  $s_j$  sont respectivement la moyenne et l'écart type de la  $j$ -ème variable dans cette même enquête,
- $z'_i = (z_i^1, \dots, z_i^q)$  est le vecteur des réponses de l'individu  $i$ , ici non centrées.

Pratique de la méthode :

■ On extrait dans **A**, avec les techniques du paragraphe (3.2.2.), le sous-ensemble des variables reconstruisant au mieux (au sens du  $R^2$ ) les  $c^\ell$ , en imposant ou excluant éventuellement certaines variables plus pertinentes pour l'étude du problème spécifique  $B^1$ .

---

<sup>1</sup> Quand on suit l'opinion (diachronie), par exemple, un grand nombre de variables communes est imposé pour des raisons d'actualité. Ici, en général :  $q > p/2$ .

Nous avons vu (3.2.2.) qu'on obtient des coefficients de corrélation multiple  $R_\rho$  de l'ordre de 0.9 ou plus, quand on régresse des  $c^\ell$  de  $A$ , se situant dans un sous-espace principal de petite dimension, par rapport au quart ou au cinquième des variables initiales. Dans la pratique, les questionnaires faisant intervenir des variables qui ne sont pas, en général, très corrélées entre elles, les composantes  $c^\ell$  ne peuvent presque jamais être des combinaisons linéaires parfaites des variables de la passerelle si  $q \ll p$ .

- Les variables "optimales" sont insérées dans le questionnaire de  $B$  accompagnées éventuellement de certaines variables du signalétique (variables qualitatives illustratives), qui serviront, entre autres, pour valider la méthode dans l'espace des individus.

- Pour obtenir la passerelle, on élimine avec les techniques du paragraphe (3.2.3.) les variables indésirables.

Dans les enquêtes diachroniques, la passerelle doit respecter au mieux les équivalences (3.5) et (3.6) ; en effet, au delà des fluctuations d'échantillonnage, une différence importante entre  $V_A$  et  $V_B$  est interprétée (car on étudie la même population) comme provenant d'un changement dans les significations.

Par contre, dans les enquêtes parallèles où les variables communes sont en principe peu nombreuses, on est moins exigeant quant au respect des conditions (3.5) et (3.6).

- On recalcule les coefficients  $b_j^\ell$  de la formule (3.13) avec les variables restantes. Pour juger de la qualité de la formule donnant la reconstruction  $\hat{c}^\ell$  de  $c^\ell$  dans  $A$ , au delà de la valeur de  $R^2$ , on peut faire comme habituellement en régression : tests  $t$  et  $F$ , erreur de prédiction moyenne, écart-type des coefficients... On tiendra compte de plus des remarques suivantes :

$$- \text{cor}(z^j, \hat{c}^\ell) = \frac{z^j{}' D_p \hat{c}^\ell}{\|z^j\| \|\hat{c}^\ell\|} = \frac{z^j{}' D_p c^\ell}{\|z^j\| \|R_\rho c^\ell\|} = \frac{\text{cor}(z^j, c^\ell)}{R_\rho} \quad \begin{array}{l} \text{pour les} \\ \text{variables } z^j \\ \text{de la passerelle,} \end{array}$$

$$- \text{cor}(y^j, \hat{c}^\ell) < \text{cor}(y^j, c^\ell), \quad \text{pour les autres variables } y^j, \text{ en général.}$$

- si les variables de la passerelle ne sont pas très corrélées entre elles, pour des valeurs élevées  $R_\ell$  des coefficients de corrélation multiple, les composantes  $\hat{c}^\ell$  seront quasiment orthogonales entre elles (argument de transitivité).

- l'inertie expliquée par les  $\hat{c}^\ell$  n'est pas comparable à l'inertie expliquée par les  $c^\ell$ .

- un problème d'interprétation peut se poser quand on passe de la détermination des  $c^\ell = \sum_{j=1}^p u_j^\ell x_j$  par l'ACP au calcul des  $\hat{c}^\ell = \sum_{j=1}^q b_j^\ell z_j$  à l'aide de la régression : les coefficients homologues  $u_j^\ell$  et  $b_j^\ell$  peuvent ne pas être du même signe.

■ Par la formule de reconstruction (3.13), on envoie ("greffe") les individus et les points moyens associés à des groupes d'individus de B sur les plans principaux de A. Remarquons ici, que (3.13) peut se réécrire :

$$\hat{c}_i^\ell = \sum_{j=1}^q b_j^\ell \frac{z_i^j - m_B^j}{s_j} = \sum_{j=1}^q b_j^\ell \left[ \frac{z_i^j - m_B^j}{s_j} + \frac{m_B^j - m_B^j}{s_j} \right] = \sum_{j=1}^q b_j^\ell \frac{z_i^j - m_B^j}{s_j} + Ct(\ell) \quad (3.14)$$

où :  $m_B^j$  est la moyenne de la j-ème variable dans B et  $Ct(\ell)$  une constante qui ne dépend pas de  $z_i$ .

Appliquer (3.13) revient donc à centrer par rapport à B et à ajouter une constante qui est supposée mesurer un décalage du point moyen global de B par rapport à l'origine qui est le point moyen de A. Cette constante doit être interprétée avec circonspection :

- dans les enquêtes parallèles, il faut tenir compte, en particulier, de l'influence des techniques d'interview qui jouent, on le sait, sur les moyennes.

- dans les enquêtes diachroniques, en principe, cette constante représente une évolution du point moyen de la population de référence. Mais, la façon de construire la passerelle (grand nombre de variables communes ; élimination de plusieurs variables pour respecter au mieux  $V_A = V_B$ ) peut conduire à des variables insuffisamment diversifiées ; on ne peut alors considérer que la quantité  $Ct(\ell) = \sum_{j=1}^q \text{Const}_j^\ell$  soit une mesure crédible de l'évolution du point

moyen global. Le problème de la représentativité des variables de la passerelle, quant à l'estimation de l'évolution globale, est pratiquement impossible ; on le résout, empiriquement, en partie par une étude de robustesse de  $Ct(\ell)$  en essayant différents jeux de variables.

■ Dans l'espace des variables de B, les composantes  $c_B^\ell$  obtenues par (3.13), expressions duales des axes homologues dans l'espace des individus, servent de cadre pour l'analyse de toutes les variables de B (utilisation classique des cercles de corrélation par la procédure des points supplémentaires).

Si  $V_A = V_B$ , les  $c_B^\ell$  sont assimilés aux composantes de A ; toutes les variables de B peuvent donc ainsi être "greffées" dans A.

Si  $V_A$  est significativement différent de  $V_B$ , il est fort probable que les mêmes combinaisons linéaires des variables de la passerelle dans A et B conduisent à des composantes non équivalentes. Dans ce cas on pourra, malgré tout, accorder aux indicateurs de B une interprétation plus au moins similaire à celle des indicateurs de référence en fonction :

- (a) de la similitude (entre A et B) des profils des corrélations entre variables de la passerelle et composantes.
- (b) de la similitude des positions des points moyens associés aux mêmes groupes d'individus sur les axes factoriels correspondants.
- (c) des angles entre composantes homologues (y a-t-il conservation de l'orthogonalité ?).

Rappelons que, d'après le principe de dualité, l'interprétation des indicateurs factoriels doit tenir compte à la fois de (a) et de (b).

#### Deux remarques importantes

■ En ce qui concerne les différents indices statistiques que l'on peut associer à la formule de reconstruction (3.13), rappelons la formule permettant d'estimer la variance  $s_j^\ell$  de l'estimateur  $b_j^\ell$  :

$$s_j^\ell \cdot \text{var}(b_j^\ell) = \frac{1}{n-q} \frac{1-R^2(c^\ell ; z^1, \dots, z^q)}{1-R^2(z^j ; (z^i ; i \neq j))}$$

Le problème de colinéarité ne se posant pas pour les variables  $z^j$ , il apparaît que les écarts-type des coefficients de régression seront petits<sup>1</sup>. Aussi, les tests t associés, seront significatifs excepté pour des coefficients  $b_j^l$  très petits en valeur absolue qui, de toute façon, contribuent peu dans le calcul des  $\hat{c}_i^l$  (voir figure 5.4.b).

■ A propos de l'examen des résidus, notons d'abord que, pour des tailles d'échantillon importantes (dans une enquête de référence, en principe  $n > 1000$ ), la contribution d'une observation aux coefficients de régression est petite. De plus, et c'est la raison pour laquelle on n'examine plus les résidus, on a toujours constaté une distribution des résidus adéquate (voir exemple dans les figures 3.3.a et 3.3.b).

Pour des valeurs de  $R_l^2$  élevées, par ailleurs, soulignons que la reconstruction de la coordonnée  $c_g^l$  d'un point moyen  $g$  associé à l'une des modalités d'une variable qualitative, liée de façon significative à la  $l$ -ème dimension factorielle, est très satisfaisante dès que l'effectif  $n_g$  du groupe correspondant est élevé.

En effet, on sait d'abord que la variance du résidu associé au point moyen est faible, puisqu'elle est  $1/n_g$  fois plus faible que celle du résidu  $e^l$  restreint aux individus du groupe. On sait de plus que le résidu  $e^l$  est corrélé ( $\text{cor}(c^l, e^l) = (1 - R_l^2)^{1/2}$ ) à la composante  $c^l$ . Si la liaison entre la composante factorielle et la variable qualitative est forte, le résidu  $e_g^l$ , moyenne "intra  $g$ " des  $e_i^l$ , est lié à la coordonnée  $c_g^l$  presque linéairement :  $e_g^l \approx \text{cor}^2(c^l, e^l) c_g^l = \alpha c_g^l$ .

On a donc :  $\hat{c}_g^l \approx (1 - \alpha) c_g^l$ . Le système des coordonnées des centres de gravité, tel qu'il est reconstruit, se déduit par contraction du système des vraies coordonnées. On voit ainsi que lors du passage de  $C$  à  $\hat{C}$  (cf. fin du paragraphe 3.2.2.), la perte d'information utile dans l'espace des individus (la disposition des centres de gravité) est petite (voir figure 3.3.c).

---

<sup>1</sup>Dans toutes les enquêtes "Agoramétrie" et pratiquement sans exception, pour la totalité des variables de l'ACP :  $R^2(x^j ; (x^i / i \neq j)) < 0.5$  ; de plus, avec une procédure de sélection comme celles du paragraphe (3.2.2.), on obtient des  $R^2(z^j ; (z^i / i \neq j))$  sensiblement inférieurs.

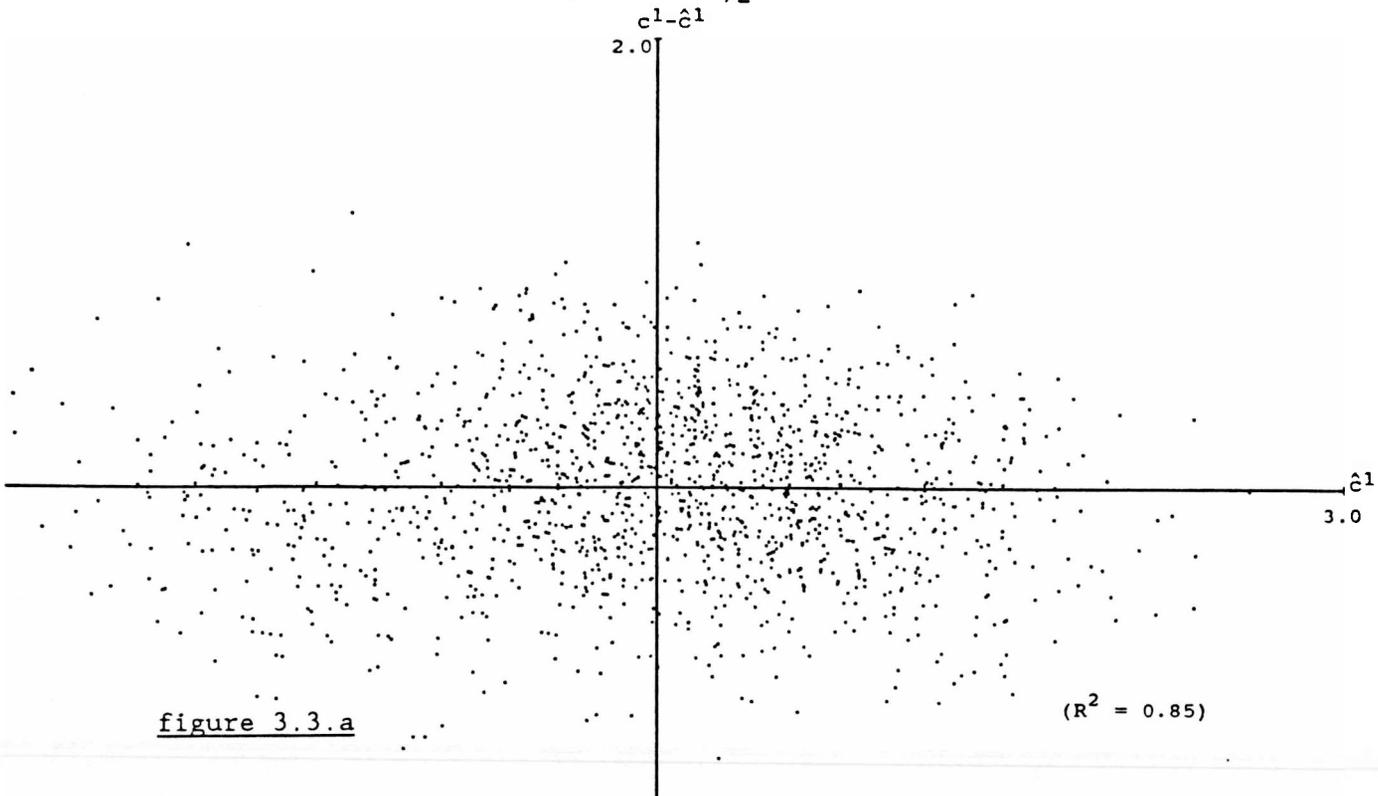


figure 3.3.a

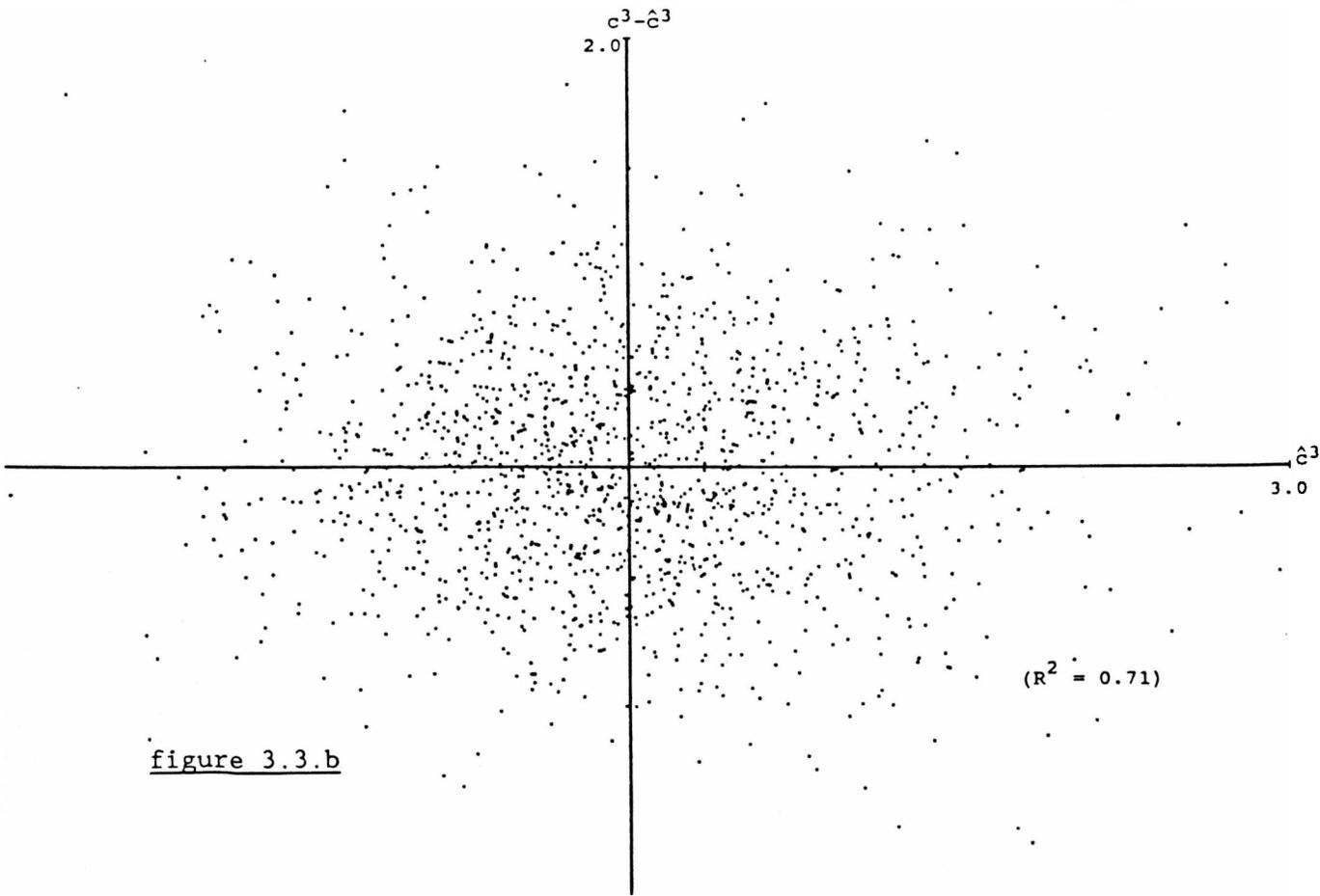


figure 3.3.b

FIGURE 3.3 : Examen des résidus de régression (Suite...)

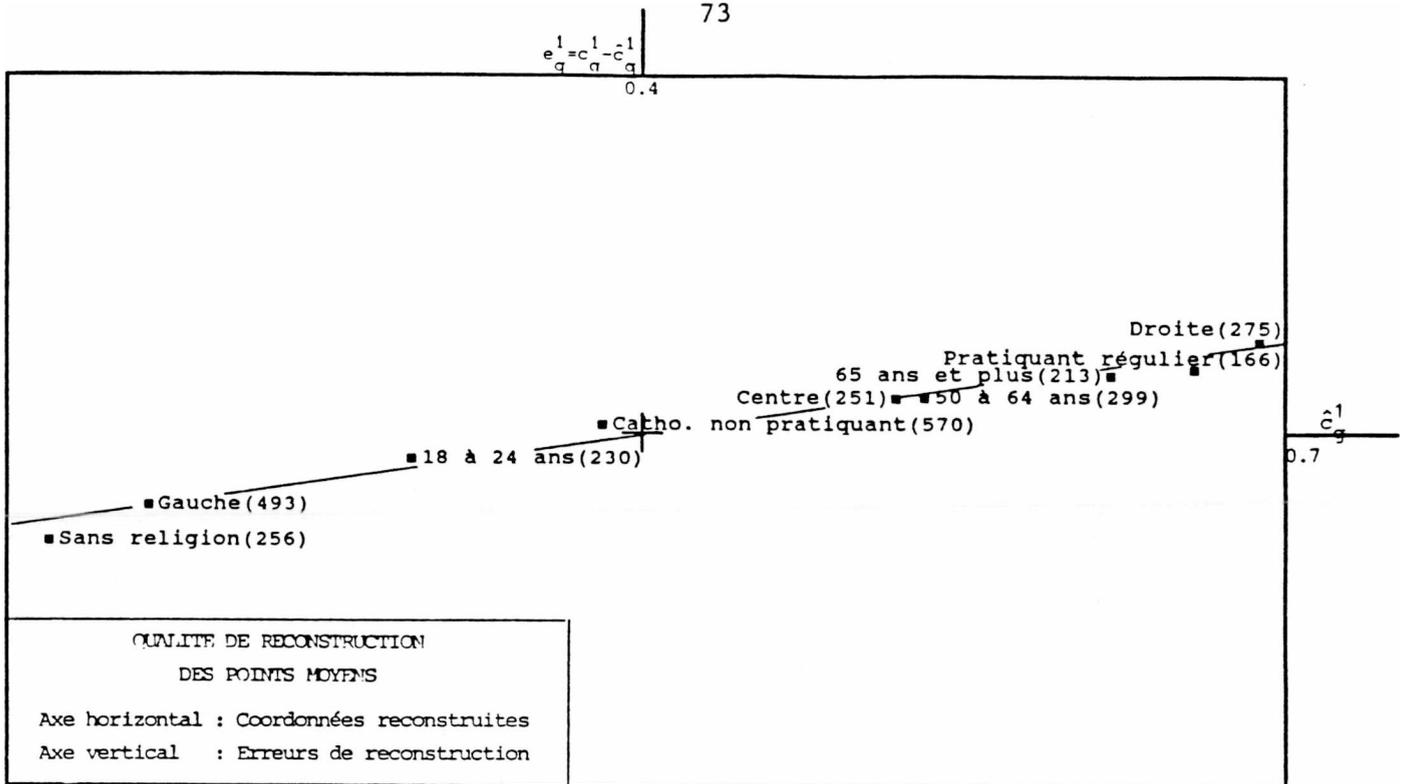


figure 3.3.c

### Régression des composantes principales par rapport aux variables de la passerelle

Enquête "Agoramétrie 1986" :

- nombre d'individus :  $n=1428$
- nombre de variables dans l'ACP :  $p=100$
- nombre de variables (optimales) de la passerelle :  $q=20$

(les composantes initiales sont normées à 1).

■ figures 3.3.a et 3.3.b : examen classique de la forme de la distribution des résidus [Draper et Smith 1981, ch. 3]. Pour la régularité de la distribution, les résidus doivent former une "bande horizontale".

■ figure 3.3.c : étude de la liaison entre les approximations fournies par la régression et les résidus pour certains points moyens caractéristiques de l'axe 1. Pour les points significatifs (effectif élevé), la forme de la liaison paraît presque linéaire.

La technique décrite ici n'est manifestement pas adaptée à une analyse fine des différences entre  $V_A$  et  $V_B$ , ce qui constitue son défaut majeur. On ne peut faire communiquer A et B qu'en approchant les composantes de A par régression sur un sous-ensemble de variables pour lequel (3.5) est vraie. Procéder ainsi, entraîne, entre autres, qu'il est assez difficile d'effectuer des greffes "par transitivité", les composantes factorielles devant être toujours des combinaisons linéaires des mêmes variables. Cette difficulté est gênante dans les enquêtes diachroniques, par exemple, où on est obligé de conserver un nombre élevé de variables communes qui ont souvent perdu, au cours du temps, de leur intérêt.

Comparée aux autres techniques possibles, l'avantage principal de la méthode de la greffe réside dans le fait qu'elle est la seule qui permet une véritable "greffe" des individus de B sur les plans de référence de A, et cela sous des conditions assez générales. Il faut par ailleurs remarquer que, dans la pratique, la condition restrictive  $V_A = V_B$  est souvent respectée :

- dans les enquêtes diachroniques, l'expérience montre qu'on arrive toujours à isoler un sous ensemble de variables respectant cette condition et permettant une assez bonne approximation des indicateurs de référence.
- dans les enquêtes parallèles effectuées sur une même population, théoriquement la condition doit toujours être vérifiée ; on dispose donc, avec l'enquête de référence, d'un moyen pour juger de la qualité de l'enquête B.
- dans les enquêtes parallèles la condition est encore en général vérifiée (en ce sens qu'il n'y a pas de contradictions apparentes entre les covariances homologues) lorsque la population de B est une sous-population de A. On a constaté, en effet, que la différenciation à l'intérieur d'un groupe particulier s'opère toujours à peu près de la même façon que dans la population de référence.

La dernière remarque, qui est d'une importance capitale pour la méthode de la greffe, est illustrée dans le graphique de la figure 5.6 (voir exemple de la section 5.2.). Dans cet exemple, on peut par ailleurs suivre les principales opérations effectuées lors de la greffe, sur le référent "Agoramétrie", d'une enquête portant sur l'action militante et la protection sociale.

### 3.2.5 Technique par rotations procrustéennes

Si les composantes factorielles de référence de **A** sont les dimensions caractéristiques du domaine étudié, elles doivent être situées dans un sous-espace principal de petite dimension de **B** contenant l'essentiel de l'information, le reste étant considéré comme du "bruit". Pour les identifier dans **B**, par l'intermédiaire des corrélations entre composantes et variables communes, on s'appuie sur des techniques de rotation procrustéenne.

Soit :

- $X_A(p_A \times n_A)$  et  $X_B(p_B \times n_B)$ , les tableaux des données centrés-réduits dans **A** et **B** respectivement,
- $Z_A(q \times n_A)$  et  $Z_B(q \times n_B)$ , les sous-tableaux correspondant aux variables de la passerelle,
- $C_A(k_A \times n_A)$ , le tableau des  $k_A$  composantes de référence normées de **A**,
- $C_B(k_B \times n_B)$ ,  $k_B \geq k_A$ , le tableau des  $k_B$  premières composantes principales orthonormées issues de l'ACP de toutes les variables de **B**.

Pour la simplicité des notations, soit encore  $A(q \times k_A) = Z_A D_p C_A'$  et  $B(q \times k_B) = Z_B D_p C_B'$  les matrices des corrélations entre variables communes et composantes.

On cherche dans l'espace factoriel principal de dimension  $k_B$  de **B** un ensemble de  $k_A$  composantes  $f^1, \dots, f^{k_A}$  normées (combinaisons linéaires des composantes principales) de tableau associé  $C_F(k_A \times n_B)$ , tel que les corrélations entre composantes  $f^i$  et variables communes dans **B** soient aussi voisines que possible des corrélations entre les composantes de référence et ces mêmes variables dans **A** ; matriciellement, cela se traduit par la proximité entre **A** et  $F = Z_B D_p C_F'$ .

Notons  $T(k_B \times k_A)$  la matrice de transformation (rotation) associée à la transposée de l'application linéaire  $R \xrightarrow[k_B]{k_A} R^{k_A}$ , telle que :  $C_F = T' C_B$ . On cherche alors  $T$ , sous la contrainte  $\text{diag}(T'T) = I_{k_A}$ , de façon à optimiser un critère qui mesure la proximité entre **A** (matrice cible) et  $F = BT$ .

Dans la pratique on utilise le plus souvent l'un des trois critères suivants:

$$\blacksquare \text{ critère procruste : } f(T) = \text{trace}(A-BT)'(A-BT) \quad (3.15)$$

- critère du produit scalaire :  $g(T) = \text{trace}(A'BT)$  (3.16)

- critère de congruence :

$$h(T) = \text{trace}\{[\text{diag}(A'A)]^{-\frac{1}{2}}A'BT[\text{diag}(T'B'BT)]^{-\frac{1}{2}}\} \quad (3.17)$$

Le critère procruste est donc un critère de moindres carrés. Le critère de congruence est défini à partir du coefficient de congruence entre deux vecteurs introduit par Tucker :

$$\phi(x,y) = \frac{x'y}{(x'x.y'y)^{\frac{1}{2}}} \quad (3.18)$$

Le critère du produit scalaire peut être vu comme un critère de congruence pondéré, où chaque congruence est pondérée par la moyenne géométrique des carrés des "longueurs" des deux vecteurs respectifs. On verra plus loin que dans le cas où T est orthogonal ce critère est équivalent au critère procruste.

On trouvera d'autres indices de proximité dans [Evans 1971] et dans [Ramsay et al. 1984].

Il ne faut pas perdre de vue que l'on traite, ici, des rotations procrustéennes dans l'espace des variables et non pas dans celui des individus ; même si les deux types de rotation acceptent une formulation algébrique presque identique, ils sont de nature différente.

Quand on travaille dans l'espace des individus, on cherche l'isométrie (translation plus rotation) à appliquer à un nuage de points-individus pour le rapprocher autant que possible, sans le déformer, d'un autre nuage de points ; conceptuellement, c'est le nuage qui est déplacé ([Gower 1975], [ten Berge 1977], [Lafosse 1985]...).

Quand on travaille dans l'espace des variables, par contre, on ne déplace pas véritablement le nuage des points-variables ; ne parlant plus de translation mais seulement de rotation, on effectue un changement de base, en passant des composantes principales aux composantes  $f^i$ , de façon à optimiser l'un des trois critères cités plus haut ([Evans 1971], [Mulaik 1972, ch.14]...). La proximité entre profils de corrélation (et la qualité de l'ajustement final) est souvent jugée sur la base des critères subjectifs dont l'utilisation ne serait pas justifiée dans l'autre type de rotation. On est aussi amené à considérer des transformations linéaires non-orthogonales des composantes initiales, qui ne

sont pas des rotations proprement dites (elles "déforment le nuage"), appelées "rotations obliques". L'utilisation d'une matrice de pondération M, tenant compte de l'importance relative des composantes dans le calcul des distances entre points-variables, n' a pas d'équivalent quand on fait des rotations dans l'espace des individus (cette matrice ne définit pas une métrique).

A l'origine de la technique par rotation se trouvent deux remarques concernant les repères factoriels issus des ACP des enquêtes annuelles sur les structures de l'opinion (diachronie) :

- les groupes sociaux occupent sur les axes principaux des différentes enquêtes des positions qui, à quelques exceptions près, peuvent être considérées comme fixes,
- les coordonnées des variables (thèmes de conflit) par rapport aux composantes principales correspondantes évoluent, parfois de façon considérable, d'une année à l'autre.

La première remarque s'explique par le fait que, dans une société structurée, les positions relatives occupées par les individus et les groupes sociaux, les statuts tels qu'ils sont perçus et exprimés, ne peuvent pas évoluer de façon significative dans une brève période. Les coordonnées des groupes sur les axes de l'ACP (axes de différenciation maximale, sur lesquels on affirme sa différence par rapport à autrui) doivent donc être relativement stables. Par contre, les composantes principales, dimensions caractéristiques autour desquelles le débat s'organise, sont sensibles aux faits majeurs<sup>1</sup> qui conditionnent de façon conjoncturelle le débat social en introduisant des thèmes nouveaux dans ce débat ou en modifiant la signification ou le poids des anciens thèmes. Toutefois, la perturbation n'est jamais énorme : on peut admettre que les composantes restent situées dans un espace "fondamental" caractéristique du système social à un état d'équilibre<sup>2</sup>.

---

<sup>1</sup> comme, par exemple, le changement de majorité après les élections en 1981 ; une rupture importante dans les places occupées par les acteurs a introduit une modification dans le jeu de l'offre et la demande des rôles (cf section 1.4).

<sup>2</sup> comme c'était la société française d'avant Mars 1981, après vingt ans de pouvoir gaulliste.

Inversement, les dimensions fondamentales doivent à leur tour se situer dans un sous-espace principal de petite dimension de chaque enquête. On pourra donc les reconstruire, à l'intérieur d'une enquête particulière, en tant que combinaisons linéaires des premières composantes principales. C'est l'objectif des techniques de rotation développées ici. Notons que l'on ne peut parler de repères strictement équivalents si on n'arrive pas à obtenir simultanément des axes et des composantes factorielles identiques.

Opérer de cette façon, en diachronie, présente un avantage supplémentaire : les composantes factorielles de chaque enquête sont extraites à partir de variables communes et non communes qui forment toujours un ensemble représentatif d'un même univers. Avec une technique de rotation — ce que l'on ne savait pas faire en régression — on peut obtenir dans les différentes enquêtes des composantes équivalentes, sans que celles-ci soient des combinaisons linéaires des mêmes variables ; en effet, une comparaison entre composantes peut s'effectuer à partir des seules coordonnées des variables communes.

Plus généralement, l'ensemble des techniques de rotation (orthogonales, obliques, avec coefficients de congruence...) que nous avons organisées en une procédure, permet facilement la mise à l'épreuve de diverses hypothèses structurelles (aussi bien en diachronie qu'en synchronie). Quoique s'appuyant sur des principes différents de ceux auxquels fait référence Jöreskog (cf paragraphe 3.1.2.), nous pouvons considérer que cette procédure relève de l'Analyse Factorielle Confirmatoire. Avant de la présenter, nous donnons d'abord les solutions de quelques problèmes de rotation qui sont les plus souvent rencontrés.

#### I. Rotations orthogonales : $T'T = I$

$$\bullet \bullet \min_T f_1(T) = \text{trace}(A-BT)'(A-BT) \quad ; \quad T'T = TT' = I_{k_A} .$$

La première solution rigoureuse à ce problème est due à [Schönemann 1966]. On la reprend ici, en procédant différemment, comme illustration d'un théorème de Kristof, généralisé par ten Berge [ten Berge 1983]. Il s'agit d'un théorème simple qui permet de trouver directement la

solution de plusieurs problèmes d'optimisation faisant intervenir des traces de produits matriciels (sans faire appel aux différenciations partielles avec les multiplicateurs de Lagrange).

Considérons deux ensembles de  $n$  matrices :  $U_1, \dots, U_n$  de rangs respectifs  $r_1, \dots, r_n$ , qui sont des sous-matrices de matrices orthogonales, et  $D_1, \dots, D_n$  qui sont elles mêmes diagonales. Le théorème de Kristof-ten Berge permet de borner la trace du produit matriciel  $U_1 D_1 U_2 D_2 \dots U_n D_n$ .

Notons  $m$  le plus grand des nombres des lignes ou des colonnes de toutes les matrices  $U_i$ . A chaque  $D_i$  faisons correspondre la matrice diagonale  $\Delta_i (m \times m)$  qui contient dans la diagonale les valeurs absolues des éléments de  $D_i$  rangées par ordre décroissant et zéro ailleurs. On a :

**Théorème de Kristof - Ten Berge**

$$-\text{trace} \Delta_1 \Delta_2 \dots \Delta_n E_r \leq \text{trace} U_1 D_1 U_2 D_2 \dots U_n D_n \leq \text{trace} \Delta_1 \Delta_2 \dots \Delta_n E_r \quad (3.19)$$

où  $r$  étant le minimum des rangs  $r_i$ ,  $E_r$  est la  $(m \times m)$  matrice qui contient  $I_r$  dans sa partie supérieure gauche et 0 ailleurs.

Démonstration : voir [Ten Berge 1983]

On a ici :

$$\begin{aligned} f_1(T) &= \text{trace}(A-BT)'(A-BT) = \text{trace}(A'A) - 2\text{trace}(A'BT) + \text{trace}(T'B'BT) \\ &= \text{trace}(A'A) + \text{trace}(B'BTT') - 2\text{trace}(A'BT). \end{aligned}$$

avec la contrainte :  $TT' = I$ , (3.20)

d'où :

$$\min_T f_1(T) \Leftrightarrow \max_T g(T) = \text{trace}(A'BT) \quad (3.21)$$

Considérons la décomposition en valeurs singulières de  $A'B$  (de rang  $r$ ) :

$$A'B = UDV' \quad (3.22)$$

Dans cette décomposition :

$$U'U = V'V = VV' = I_{k_A},$$

et  $D$  est la matrice diagonale des valeurs singulières ( $r$  premiers termes positifs,  $(k_A - r)$  derniers termes nuls).

On a alors :

$$g(T) = \text{trace}(UDV'T) \quad (3.23)$$

Les matrices  $V$  et  $T$  étant orthogonales,  $V'T$  l'est aussi. En appliquant le théorème de Kristof-ten Berge on obtient :

$$-\text{trace } D \leq \text{trace}[(U)D(V'T)] \leq \text{trace } D \quad (3.24)$$

le maximum de  $g(T)$  est alors atteint pour

$$T = VU' \quad (3.25)$$

Il est clair par (3.21) que les problèmes (3.15) et (3.16) coïncident quand  $k_A = k_B$ . Pour  $k_B > k_A$ , (3.25) reste solution au problème (3.16),  $T$  vérifiant maintenant seulement ( $T'T = I$ ) ; par contre, (3.20) n'étant plus vérifiée, (3.25) n'est pas solution au problème procruste (3.15).

$$\blacksquare \min_T f_2(T) = \text{trace}(A-BT)'(A-BT) ; T'T = I_{k_A} ; k_B > k_A$$

Pour ce problème, il n'y a pas pour l'instant de solution analytique, sauf pour le cas  $k_A = 1$  (on se ramène alors à un problème de rotation "oblique" (voir minimisation de la fonction  $f_6$ , plus loin)). On peut cependant obtenir une solution par itérations comme suit :

**Etape 0** : Transformer la  $(q \times k_A)$  matrice  $A$  en une matrice  $(q \times k_B)$   $A^*$  en rajoutant  $k_B - k_A$  colonnes arbitraires.

————— Itérer sur les deux étapes suivantes —————

**Etape 1** : Résoudre le problème procruste de "plein rang" ( $k_A = k_B$ ) pour les matrices  $A^*$  et  $B$  ; soit  $T^*(k_B \times k_B)$  la matrice de rotation obtenue.

**Etape 2** : Partitionner  $T^*$  en  $T^* = [T | T_\star]$ , où  $T$  est de dimension  $(k_B \times k_A)$  et  $T_\star$  de dimension  $(k_B \times (k_B - k_A))$  ; remplacer alors les  $(k_B - k_A)$  dernières colonnes de  $A^*$  par  $BT_\star$ .

Pour s'assurer d'avoir évité des minima locaux, on devra répéter la procédure plusieurs fois avec différents choix initiaux pour les  $k_B - k_A$  dernières colonnes de  $A^*$ . Selon Ten Berge et Knol [Ten Berge et Knol 1984], on peut éviter ces répétitions en choisissant pour  $T_\star^0$  une matrice constituée des  $k_B - k_A$  derniers vecteurs propres de  $B'B$  et en posant  $A_0^* = [A | BT_\star^0]$ .

En extrapolant ce raisonnement, on pourra donner une solution itérative même quand il y a des éléments inconnus dans la matrice cible  $A$  ; on remplace à chaque itération l'élément manquant  $a_{ij}$  par l'élément homologue de la matrice  $BT^*$ .

$$\bullet \bullet \max_T h_1(T) = \text{trace}([\text{diag}(A'A)]^{-\frac{1}{2}} A'BT[\text{diag}(T'B'BT)]^{-\frac{1}{2}}) ; T'T - TT' = I$$

Il s'agit de maximiser la somme des coefficients de congruence entre profils homologues des corrélations.

Du point de vue de l'interprétation, deux composantes factorielles sont considérées comme équivalentes quand les profils des corrélations sont identiques ou proportionnels. L'indice  $\phi$  (3.18) est alors un excellent indice pour mesurer la similitude entre profils des corrélations. On constate, en particulier, une bonne adéquation entre l'utilisation de  $\phi$  et le jugement subjectif sur lequel repose l'interprétation des composantes factorielles (Broken[1983]).

La solution au problème d'optimisation a été donnée par Broken [Broken 1983 et 1985], mais son algorithme, qui fait appel à la minimisation numérique, coûte cher en temps de calcul et en place mémoire. Par ailleurs, pour des problèmes de "plein rang" ( $k_B = k_A$ ), Broken a trouvé que le critère procruste (3.15) conduit à des solutions presque aussi bonnes au sens de  $\phi$  que le critère (3.17).

Pour ces raisons, et pour pouvoir traiter de l'équivalence au sens de profils de corrélation, nous proposons, à la place de la maximisation de  $h_1(T)$ , la minimisation suivante d'une fonction  $f_3$  :

$$\bullet \bullet \min_{T,D} f_3(T,D) = \text{trace}[(AD-BT)'(AD-BT)]$$

où D est une matrice diagonale de coefficients de "dilatation" et  $T'T - TT' = I_{k_A}$ .

Schönemann et Carroll [Schönemann et Carroll 1970] ont donné la solution analytique à un problème similaire en considérant un seul coefficient de dilatation global d. On propose ici une solution itérative pour la minimisation de  $f_3(T,D)$ .

Rappelons d'abord, en utilisant les notations classiques, les résultats suivants pour la dérivation des traces des matrices :

$$\frac{\partial \text{tr}(AX)}{\partial X} = A' \quad ; \quad \frac{\partial \text{tr}A}{\partial X'} = \left[ \frac{\partial \text{tr}A}{\partial X} \right]'$$

$$\frac{\partial \text{trg}}{\partial (\text{diag} X)} = \text{diag} \left( \frac{\partial \text{trg}}{\partial X} \right) \quad ; \quad \frac{\partial \text{tr}(X'AX)}{\partial X} = (A+A')X$$

où  $A$  et  $X$  sont des matrices carrées et  $g$  une expression matricielle en  $X$

Pour  $T$  fixé, on a :

$$\frac{\partial f_3(T,D)}{\partial D} = \frac{\partial \text{trace}[D'A'AD - 2D'A'BT + T'B'BT]}{\partial D} = 0$$

ceci s'écrit :

$$\text{diag}(2A'AD) - \text{diag}(2A'BT) = 0$$

d'où, si  $A'A$  est inversible :

$$D = [\text{diag}(A'A)]^{-1} \text{diag}(A'BT) \quad (3.26)$$

Quand on considère un seul coefficient  $d$  de dilatation globale :

$$d = \text{trace}(A'BT) / \text{trace}(A'A) \quad (3.27)$$

La solution itérative au problème d'optimisation pour  $T$  et  $D$  est alors :

**Etape 0** : Poser  $D^* = I_{k_A}$

————— Itérer sur les deux étapes suivantes —————

**Etape 1** : Minimiser  $f(T, D^*)$  par rapport à  $T$ . Soit  $T^*$  la solution obtenue.

**Etape 2** : Minimiser  $f(T^*, D)$  par rapport à  $D$ . Soit  $D^*$  la solution.

Remarque : Quand on doit tenir compte à la fois des dilatations et des éléments manquants dans  $A$ , on peut combiner les itérations qu'impliquent la minimisation de  $f_2(T)$  et celle de  $f_3(T, D)$  en un seul programme, en faisant converger les deux fonctions  $f_2$  et  $f_3$  simultanément. Notre expérience montre que l'on converge rapidement vers un optimum.

**Pondération des variables et des axes :**

Pour les variables :

$$\bullet \bullet \min f_4(T) = \text{trace}[(A-BT)'D_V(A-BT)]$$

où  $D_V$  est une matrice de pondération des variables.

On vérifie immédiatement que la solution à ce problème est toujours donnée par (3.25) en remplaçant  $A'B$  par  $A'D_{\alpha}B$  dans (3.22).

Pour les axes :

$$\bullet \bullet \min_T f_5(T) = \text{trace}[(A-BT)D_a^2(A-BT)'] ; \quad T'T=TT'=I,$$

où  $D_a$  est une matrice définie positive.

Pour ce problème de pondération des axes, il n'y a pas (pour l'instant) de solution simple. Une façon d'aborder le problème est de considérer, comme dans [Lissitz et al. 1976], la minimisation de  $f_5(T)$  sous la contrainte  $T'D_a^2T = I$ . Dans ce dernier cas on vérifie aisément que la solution est toujours donnée par (3.25) où  $T$  est remplacé par  $T^*=TD_a$  et en remplaçant  $A'B$  par  $A'BD_a$  dans (3.22).

## II. Rotations obliques : $\text{diag}(T'T)=I_{k_A}$

Avec les rotations obliques, on s'affranchit de la contrainte d'orthogonalité dans la comparaison des structures factorielles. Une justification théorique aux rotations obliques a été donnée par Meredith<sup>1</sup>. En étudiant une sous-population extraite d'une population parente par "sélection" sur certaines variables continues (on ne retient alors que les individus pour lesquels les variables prennent une valeur supérieure (ou inférieure) à un seuil), il a montré que la matrice des "saturations" peut être conservée à condition d'accepter une modification de la matrice de variance entre composantes factorielles.

En ce qui nous concerne, on utilisera les rotations obliques surtout comme moyen de vérification de l'orthogonalité des composantes obtenues par des rotations orthogonales.

$$\bullet \bullet \min_T f_6(T) = \text{trace}(A-BT)'(A-BT) ; \quad \text{diag}(T'T)=I_{k_A}$$

Ce problème peut être décomposé en  $k_A$  problèmes de minimisation indépendants : déterminer pour  $j=1, \dots, k_A$ , la colonne  $t_j$  de  $T$ , sous la contrainte  $t_j't_j=1$ , qui minimise  $f_7(t_j)=(a_j-Bt_j)'(a_j-Bt_j)$ . On trouve une solution rigoureuse au problème de la minimisation de  $f_7(T)$  dans [Ten

---

<sup>1</sup>Cité dans [Mulaik 1972, ch.14 : "Comparing Factors Across Populations"].

Berge 1979]. Dans les calculs, les deux étapes les plus pénalisantes sont la diagonalisation de  $B'B$  et la maximisation d'une fonction relativement complexe par la méthode de Newton-Raphson. Les temps d'exécution sont, par conséquent, insignifiants.

$$\blacksquare \max_T h_2(T) = \text{trace} \left\{ [\text{diag}(A'A)]^{-\frac{1}{2}} A'BT [\text{diag}(T'B'BT)]^{-\frac{1}{2}} \right\} ; \text{diag}(T'T) = I$$

Il s'agit de maximiser (individuellement) les coefficients de congruence entre composantes homologues. La solution à ce problème se présente sous une forme assez simple [Korth et al 1976] :

- Calculer  $T^* = (B'B)^{-1} B'A$
- Soit  $D_T^2 = \text{diag}(T^* T^*)$
- On obtient alors  $T = T^* D_T^{-1}$  (3.28)

Comparaison entre composantes orthogonales et obliques :

$$\blacksquare \min_T g_1(T) = \text{trace} \left[ (TC-C)' D_p (TC-C) \right] ; \text{diag}(CD_p C') = I ; TCD_p C' T' = I$$

On cherche ici à obtenir un système de composantes orthonormées de tableau associé  $TC$ , les plus proches possible (prises individuellement) d'un ensemble de composantes corrélées de tableau associé  $C$ .

La solution pour  $T$  est encore assez simple [Price et al. 1977]. Soit  $R = UD_\lambda U'$  la décomposition en valeurs singulières de  $R = CD_p C'$  ; alors :

$$T = R^{-\frac{1}{2}} = UD_\lambda^{-\frac{1}{2}} U' \quad (3.29)$$

Procédure proposée :

Les calculs impliqués dans les problèmes d'optimisation évoqués sont peu coûteux ; en effet, les calculs les plus pénalisants interviennent au moment de la décomposition en valeurs singulières : diagonalisation de petites matrices symétriques (de l'ordre de  $(k_B \times k_B)$ ). La procédure que nous proposons est basée sur un logiciel conversationnel effectuant tous les calculs décrits ; en voici les grandes lignes :

■ On effectue dans B une analyse en composantes principales sur :

(i) la totalité des variables, si celles-ci peuvent être considérées comme représentatives du même univers que celles de A,

(ii) les variables communes seulement, dans le cas contraire. Ici, il est important que les variables communes permettent de reconstruire dans A de façon satisfaisante les indicateurs factoriels de référence et que de plus elles soient assez diversifiées (voir remarque du paragraphe 3.2.2.). En toute rigueur, on pourrait introduire certaines variables de B non communes, à condition qu'elles ne soient pas redondantes avec les autres variables de l'ACP et qu'elles ne perturbent pas de façon significative l'espace factoriel.

■ Avec les techniques du paragraphe 3.2.3., on repère les variables qui sont manifestement perçues de façon différente dans les deux enquêtes. Les coordonnées des variables restantes (variables de la passerelle) formeront les lignes de la matrice A. Notons que, théoriquement, pour effectuer une rotation procrustéenne dans un espace défini par m composantes, on a besoin d'une cible de dimension ( $m \times m$ ). On peut par conséquent pousser assez loin l'élimination de variables, un minimum de "redondance" étant toutefois nécessaire pour la stabilité des résultats.

■ on effectue une première série de rotations en faisant varier le nombre d'axes  $k_B$ . On commence par poser  $k_B = k_A$  et on incrémente  $k_B$  jusqu'à ce que la nouvelle ligne de la matrice de rotation T contienne uniquement des éléments petits en valeur absolue. Il est clair qu'en augmentant  $k_B$  on améliore la valeur de l'indice d'ajustement entre A et BT ; cette amélioration doit toutefois être significative.

■ Ayant choisi  $k_B$ , on effectue des rotations (orthogonales d'abord) en faisant intervenir différents sous-ensembles des variables de la passerelle. C'est en examinant la décomposition de l'écart procruste par axe et par variable, que l'on modifie à tâtons les valeurs de certains éléments de la cible (éventuellement pour réduire leur influence à zéro dans les rotations).

On compare la solution orthogonale sans dilatations à celle avec coefficients de dilatation (un coefficient par axe ou un coefficient

global), puis à celles correspondant aux rotations obliques (avec ou sans dilatations). Si on ne progresse pas ainsi dans l'ajustement de façon significative, on garde comme solution finale la solution orthogonale simple. Notons que l'utilisation des coefficients de dilatation est en harmonie avec l'utilisation de l'équivalence (3.7) entre matrices de corrélation.

Des procédures de construction d'intervalles de confiance pour les différents paramètres (éléments de BT, corrélations entre composantes), avec des techniques du type jackknife ou bootstrap, sont aussi envisageables.

Remarque : la qualité de l'ajustement global est caractérisée en fonction de l'écart carré moyen :

$$e = \left[ \frac{\text{trace}(\text{AD-BT})'(\text{AD-BT})}{N} \right]^{1/2} \quad (3.30)$$

où N est le nombre d'éléments actifs intervenant dans la rotation ( $N \leq qk_A$ ).

■ Les composantes obtenues dans l'espace des variables  $R^{n_B}$  après rotation, correspondent par dualité à des axes factoriels dans l'espace des individus  $R^{p_B}$  de B. Les coordonnées (sur ces axes) des modalités de variables qualitatives communes avec A, fournissent un moyen supplémentaire pour juger de l'équivalence entre indicateurs dans les deux enquêtes. Ici, il ne faut pas cependant perdre de vue que :

(i) une rotation orthogonale dans  $R^{n_B}$  correspond à une rotation oblique, en général, dans  $R^{p_B}$  : appliquer la matrice de rotation T' aux composantes principales  $c_B^{\ell}$  dans  $R^{n_B}$  ( $C_F = T' C_B$ ) correspond à une transformation oblique des axes principaux  $u_B$  dans  $R^{p_B}$  :

$$U_F = D_T^{-1} (T' D_{1/\lambda}) D_T^{-1} U_B$$

où,  $D_{\lambda}$  étant la matrice diagonale des valeurs propres correspondantes :

$$D_T^2 = \text{diag}(T' D_{1/\lambda} T)$$

(ii) Les échelles utilisées pour tracer les plans principaux dans A et B ne sont pas comparables ; il en est de même pour les points moyens globaux de A et de B qui, tous les deux, sont confondus avec les origines de leurs plans principaux respectifs. Par conséquent, ces plans ne sont plus superposables, alors qu'ils l'étaient avec la technique par régression.

La procédure décrite, que l'on peut suivre à travers l'exemple de la section 5.3. permet donc de tester des hypothèses très variées sur la structure factorielle de **B** en tenant compte de la structure factorielle de **A**. Son choix semble entièrement justifié pour effectuer de la communication entre enquêtes quand, et c'est le cas dans les enquêtes diachroniques, une hypothèse plausible à tester, est du type "*conservation des structures factorielles à une rotation orthogonale près*".

De plus, elle répond de façon assez satisfaisante aux remarques effectuées dans le paragraphe concernant l'approche de Jöreskog, en ce sens que :

(i) toute hypothèse est testée sur la matrice de variance observée sur l'échantillon, et non pas sur une matrice de variance théorique ; par ailleurs, la recherche d'identité structurelle s'effectue dans un sous-espace de dimension relativement petite (espace significatif) par rapport au nombre de variables initiales.

(ii) l'information apportée par les variables qualitatives, et plus généralement, la dualité individus-variables sont mieux prises en compte.

(iii) l'efficacité au niveau des calculs fait de cette procédure une méthode d'analyse factorielle confirmatoire conversationnelle pour le traitement de problèmes de taille importante.

Certes, on ne peut ici prétendre à une analyse des écarts entre matrices de variance aussi fine que celle que permet la méthodologie de Jöreskog. Compte tenu, cependant, du grand nombre de sources d'erreur de mesure et du flou relatif des variables (d'opinion) mesurées, seuls les effets majeurs (par exemple une rotation) sont ici significatifs. La procédure par rotations semble répondre, de façon plus réaliste que l'approche inférentielle de Jöreskog, à la détection de ces effets .

## BIBLIOGRAPHIE

- Bonifas, L., Y. Escoufier, P. L. Gonzalez & R. Sabatier (1984). *Choix de variables en analyse en composantes principales*. Revue de Statistique Appliquée, XXXII, 5-15.
- ten Berge, J. M. F. (1977). *Orthogonal procrustes rotation for two or more matrices*. Psychometrika, 42, 2, 267-276.
- ten Berge, J. M. F. (1983). *A generalisation of Kristof's theorem on the trace of certain matrix products*. Psychometrika, 48, 519-523.
- ten Berge, J. M. F. & D. L. Knol (1984). *Orthogonal rotation to maximal agreement for two or more matrices of different column orders*. Psychometrika, 49, 49-55.
- ten Berge, J. M. F. & K. Nevels (1977). *A general solution to Mosier's oblique procrustes problem*. Psychometrika, 42, 593-600.
- Broken, F. B. (1983). *Orthogonal procrustes rotation maximizing congruence*. Psychometrika, 48, 343-352.
- Broken, F. B. (1985). *The simultaneous maximization of congruence for two or more matrices under orthogonal rotation*. Psychometrika, 50, 51-56.
- Cailliez, F. & J.P. Pagès (1976). *Introduction à l'analyse des données*. Paris : SMASH.
- Croquette A., A. Kobilinsky, J. P. Pagès & Y. Schektman (1983). *Optimalité en analyse factorielle*. Cahiers du B.U.R.O. n°43-44, 3-41.
- Dambroise, E., Y. Escoufier & P. Massote (1987). *Application de l'Analyse de Données à l'élaboration de mini sondages d'opinion*. Revue de la Statistique Appliquée, XXXV, 1, 9-24.
- Draper, N. & H. Smith (1984). *Applied Regression Analysis*, 2nd edition. Wiley.
- Escoufier, Y. (1985). *Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données*. Statistique et Analyse de Données, 10, 1-10.
- Evans, G. T. (1971). *Transformation of factor matrices to achieve congruence*. British Journal of Mathematical and Statistical Psychology, 24, 22-48.
- Foucart, T. (1984). *Analyse Factorielle de tableaux multiples*. Paris : Masson.

- Glaçon, F. (1981). *Analyse Conjointe de plusieurs matrices de données*. Thèse de 3ème cycle, L'Université Scientifique et Médicale de Grenoble.
- Gower, J. C. (1976). *Generalised procrustes analysis*. *Psychometrika*, 40, 1, 33-51.
- Guttman, L. (1985). *The illogic of statistical inference for cumulative science*. *Applied Stochastic Models and Data Analysis*, 1, 3-10.
- Israëls, A. Z. (1986). *Interpretation of redundancy analysis : rotated vs unrotated solutions*. *Applied stochastic models and data analysis*, 2, 121-130.
- Johansson, J. K. (1981). *An extension of Wollenberg's redundancy analysis*. *Psychometrika*, 46, 93-103.
- Joliffe, I. T. (1972). *Discarding variables in a Principal Components Analysis I : artificial data*. *Applied Statistics*, 16, 225-236.
- Joliffe, I. T. (1973). *Discarding variables in a Principal Components Analysis II : real data*. *Applied statistics*, 22, 21-31.
- Joliffe, I. T. (1986). *Principal Components Analysis*. New-York : Springer-Verlag.
- Jöreskog, K. G. (1969). *A general approach to confirmatory maximum likelihood Factor Analysis*. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. (1971). *Simultaneous Factor Analysis in several populations*. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G. (1978). *Structural Analysis of Covariance and Correlation matrices*. *Psychometrika*, 43, 443-477.
- Jöreskog, K. G. & D. Sörbom (1984). *LISREL V : Analysis of linear structural relationships by the method of maximum likelihood*. Mooresville, IN : Scientific Software, Inc.
- Korth, B. & L. R. Tucker (1976). *Procrustes matching by congruence coefficients*. *Psychometrika*, 41, 531-535.
- Krishnaiah, P. R. & J. C. Lee (1980). *Likelihood ratio tests for mean vectors and covariance matrices*. IN : P.R. Krishnaiah, ed., *Handbook of statistics*, vol. 1, 513-570. North Holland.
- Krzanowski, W. J. (1987). *Selection of variables to preserve Multivariate Data Structure using Principal Components*. *Applied Statistics*, 36, 1, 22-33.
- Kshirsagar, A. M. (1972). *Multivariate analysis*. New York : Dekker.
- Lafosse, R. (1985). *Analyses procustéennes de deux tableaux*. Thèse de 3ème cycle, Université Paul Sabatier de Toulouse.
- Lee, S. Y. & K. L. Tsui (1982). *Covariance Structures Analysis in Several Populations*. *Psychometrika*, 50, 83-90.

- Lissitz, R. W., P. H. Schönemann & J. C. Lingoes (1976). A solution to the weighted procrustes problem in which the transformation is in agreement with the loss function. *Psychometrika*, 41, 547-550.
- Manly, B. F. J. & J. C. W. Rayner (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika*, 74, 4, 841-847.
- Matsueda, R. L. & W. T. Bielby (1986). *Statistical power in covariance structure models*. Washington, IN : Sociological methodology, American statistical association.
- McArdle, J. J. & R. P. McDonald (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- McCabe, G. P. (1984). *Principal variables*. *Technometrics*, 26, 137-144.
- McDonald, R. P. (1986). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 33, 161-183.
- McDonald, R. P. (1980). Describing the elephant: structure and function in multivariate data. *Psychometrika*, 51, 513-534.
- Mulaik, S. A. (1972). *The foundations of Factor Analysis*. New York : McGraw Hill.
- Mulaik, S. A. (1986). Factor Analysis and Psychometrika. Major developments. *Psychometrika*, 51, 23-33.
- Nhun Fat, T. (1984). *Problèmes posés par la mise en place d'un système de suivi des opinions. Du recueil des informations à l'utilisation des résultats*. Thèse de 3ème cycle, Université de Paris-VI.
- Price, J. M. & Nicewander, W. A. (1977). Maximally correlated orthogonal composites and oblique factor analytic solutions. *Psychometrika*, 42, 3, 439-442.
- Ramsay, J. O., J. M. F. ten Berge & G. P. H. Styan (1984). Matrix correlation. *Psychometrika*, 49, 403-423.
- Robert, P. & Y. Escoufier (1976). A unifying tool for linear multivariate statistical methods : the RV-coefficient. *Applied Statistics*, 25, 257-265.
- Satorra, A. & W. E. Saris (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50, 83-90.
- Schönemann, P. H. (1966). A generalised solution of the orthogonal procrustes problem. *Psychometrika*, 31, 1-10.
- Schönemann, P. H. & R. M. Carroll (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35, 245-255.

## CHAPITRE IV

### SUR LE TRAITEMENT INFORMATIQUE DES ENQUETES D'OPINION

#### INTRODUCTION

Pour traiter ses données le statisticien contemporain a le plus souvent recours aux grands logiciels classiques comme BMDP<sup>1</sup>, SAS<sup>2</sup>, SPSS<sup>3</sup>... Résultant d'un travail de longue haleine d'équipes spécialisées et de mises à jour régulières, ces logiciels constituent des outils indispensables, voire de véritables environnements informatiques (SAS), permettant le traitement d'une grande variété de types de données.

Cependant quels qu'en soient les mérites, ces logiciels ne peuvent répondre de façon entièrement satisfaisante aux exigences de l'analyste professionnel confronté au traitement systématique d'un même type de données. Celui-ci demande alors une plus grande souplesse au niveau des entrées-sorties, une présentation des résultats plus conforme à ses propres convictions, un certain nombre de fonctions statistiques qui ne peuvent pas être incluses dans des logiciels ayant un caractère trop général... Ainsi, on a vu se développer, notamment sur micro-ordinateur, une autre génération de logiciels. Spécialisés, plus complets dans leur domaine d'application relativement restreint (régression, classification, ...), les plus récents parmi eux disposent par ailleurs d'inter-

---

<sup>1</sup>Biomedical Computer Programs, P-series

<sup>2</sup>Statistical Analysis System

<sup>3</sup>Statistical Package for the Social Sciences

faces adéquates avec les grands logiciels cités et aussi avec d'autres logiciels professionnels tels que les tableurs ou les systèmes de gestion de bases de données.

L'absence de logiciels, efficaces à nos yeux, pour le traitement d'enquêtes d'opinion, nous a conduit à élaborer, sur micro-ordinateur, un outil informatique reflétant nos idées dans ce domaine. L'ensemble des procédures mises en place devrait permettre le dépouillement, au moindre coût, des nombreuses grosses enquêtes que nous traitons annuellement, et contribuer à la promotion de notre système particulier de suivi des opinions. Trois principes directeurs nous ont alors guidé ; il fallait que le logiciel puisse :

- *traiter rapidement des fichiers de données de taille importante,*
- *produire des sorties imprimante peu volumineuses et de grande qualité,*
- *inciter l'utilisateur à concevoir d'une certaine façon les enquêtes et à effectuer sur les données un certain type de traitement statistique.*

■ L'efficacité au niveau des calculs est cruciale. Dans le contexte de la micro-informatique (contact direct avec l'ordinateur et les périphériques), il s'agissait de faire en sorte que les temps de traitement, pour des tableaux de données qui peuvent faire intervenir des milliers d'individus et des centaines de variables, deviennent, avec une programmation adéquate, comparables, sinon inférieurs, à ceux obtenus sur des gros systèmes.

Nous avons alors été initialement confronté au problème du choix du langage de programmation qui, il faut le souligner, avec la montée en puissance des ordinateurs et l'évolution des compilateurs, n'est plus aujourd'hui si important. Ayant d'abord opté pour le Pascal, nous nous sommes par la suite tourné vers le langage C, dès qu'un environnement de développement satisfaisant était disponible. Les deux langages permettent, grâce aux structures des données dynamiques, une gestion parfaite de la mémoire centrale. Cette caractéristique est particulièrement souhaitable pour un logiciel d'Analyse des Données dans la mesure où on peut, par exemple, se "débarasser" facilement de certaines limites

fatidiques de dimensionnement : nombre maximum de variables, nombre de modalités par variable... La seule limite reste, évidemment, la taille de la mémoire disponible.

Le langage C, en particulier, est plus approprié pour le développement des programmes modulaires et portables ; il offre à l'analyste expérimenté des possibilités de programmation interdites par la syntaxe restreignante des langages fortement "typés" comme le Pascal. Enfin, dans la plupart des systèmes informatiques, on en trouve des compilateurs qui génèrent un code extrêmement efficace.

Le choix des algorithmes, pour certaines procédures intensément utilisées en Analyse de Données (inversion ou diagonalisation de matrices symétriques, tri, ...), n'est pas d'une importance capitale non plus. Sans chercher la perfection, nous avons retenu des algorithmes classiques, bien testés et assez précis et efficaces. On trouvera, outre des commentaires, des indications sur quelques temps d'exécution caractéristiques dans les sections suivantes.

Pour atteindre les performances recherchées, le choix décisif a été celui concernant l'organisation des entrées-sorties. Tout a été conçu pour réduire au maximum :

- le temps d'introduction des différents paramètres, au détriment parfois de la convivialité. Nous avons ainsi préféré rester très opérationnel pour le spécialiste, plutôt que de donner au logiciel un caractère plus "grand public",
- le volume des sorties imprimante. Dans l'environnement micro-informatique l'impression des résultats reste une partie très pénalisante pour la durée des traitements,
- le nombre d'appels des programmes concernant des tâches répétitives, par la construction systématique de modules "à un seul passage". Le module pour les tris croisés, par exemple, permet la construction et l'impression, en un seul passage, de plusieurs milliers de tableaux provenant des croisements entre deux groupes de variables (voir section 4.3.).

■ La réduction du volume des sorties ne vise pas seulement à améliorer la durée des traitements, mais aussi à faciliter le travail

d'interprétation de l'analyste qui, submergé par un nombre parfois impressionnant de pages de listings, se trouve obligé de produire un effort considérable pour en extraire l'information utile. Avec la majorité des logiciels, par ailleurs, un effort de secrétariat supplémentaire doit être produit pour créer, à partir des sorties initiales de mauvaise qualité, un document présentable.

C'est bien à ces niveaux que nous avons prioritairement cherché à nous démarquer. Tirant profit des possibilités offertes par les périphériques modernes, tels les imprimantes Laser et les tables traçantes, nous nous sommes efforcé à :

- remplacer les sorties volumineuses par des graphiques parlants et par des tableaux-résumé "intelligents",
- fournir des sorties très soignées de qualité "imprimerie", utilisables directement dans un rapport final.

Pour permettre la production de documents personnalisés, nous avons implanté un certain nombre de procédures d'entrée-sortie générales qui offrent à l'utilisateur des possibilités habituellement réservées aux logiciels d'édition : formatage des lignes et des pages, choix de polices des caractères... Divers "drivers" (pas trop généraux, car lents) assurent la compatibilité avec les principaux organes périphériques du marché.

■ Les modules sélectionnés pour les analyses statistiques, de même que leur ordre d'enchaînement, reflètent bien notre façon d'envisager la mise au point et le traitement des enquêtes d'opinion ; on incite l'utilisateur à adopter une même approche. Grâce aux analyses et aux options particulières proposées, et à un regroupement des procédures quelque peu inhabituel, celui-ci est conduit à adopter, au moins en partie, notre séquence de traitements. Donnons deux exemples :

- la qualité des sorties pour les tris à plat et les tris croisés, dans le cas général, est meilleure quand on dispose de paquets de variables ordinales ayant les mêmes modalités (exemple nos "thèmes"). Cette homogénéité autorise le recours aux coefficients de corrélation que l'on encourage par une présentation intéressante de la matrice des corrélations. Incité à construire son questionnaire en conséquence, l'utilisa-

teur est alors conduit logiquement à adopter l'ACP comme technique d'analyse multidimensionnelle.

- le calcul des coordonnées sur les axes factoriels des modalités des variables qualitatives illustratives est effectué dans un module bien séparé du module principal de l'ACP (ou de l'AFC) ; sauf demande explicite, ce calcul est accompagné des analyses de la variance permettant d'apprécier les relations qu'entretiennent chacune des composantes factorielles avec chacune des variables qualitatives. L'utilisateur en tiendra certainement compte pour interpréter les axes factoriels et pour décider du nombre d'axes à retenir (voir section 4.4.3.). Il découvre en même temps, s'il y a lieu, l'intérêt de disposer d'un signalétique "riche".

Développé sur les bases précédentes, l'outil a été construit de façon modulaire, autour d'un module principal contenant des procédures générales d'entrées-sorties et les principaux algorithmes. Dans l'organigramme de la chaîne des traitements de la figure 4.1, on peut distinguer trois catégories de modules : ceux qui permettent la création du fichier de travail à partir des données brutes (éditeur de gros fichiers, recodage, redressement... (trait simple)) ; ceux qui sont relatifs aux traitements statistiques standards (tris à plat, tris croisés, ACP, AFC, classification, communication entre enquêtes... (trait simple sur fond grisé)) ; enfin ceux qui, venant s'ajouter progressivement, concernent des traitements particuliers (rotations analytiques, classification des variables, étude de la stabilité des points moyens (trait double)).

Dans les sections suivantes on trouve une description rapide des principaux modules (sauf ceux pour la communication entre enquêtes qui ont été commentés dans le troisième chapitre). Nous nous sommes restreint aux modules qui présentent un intérêt théorique (rappel alors effectué) ou méthodologique particulier, ou encore une originalité au niveau de la présentation des résultats.

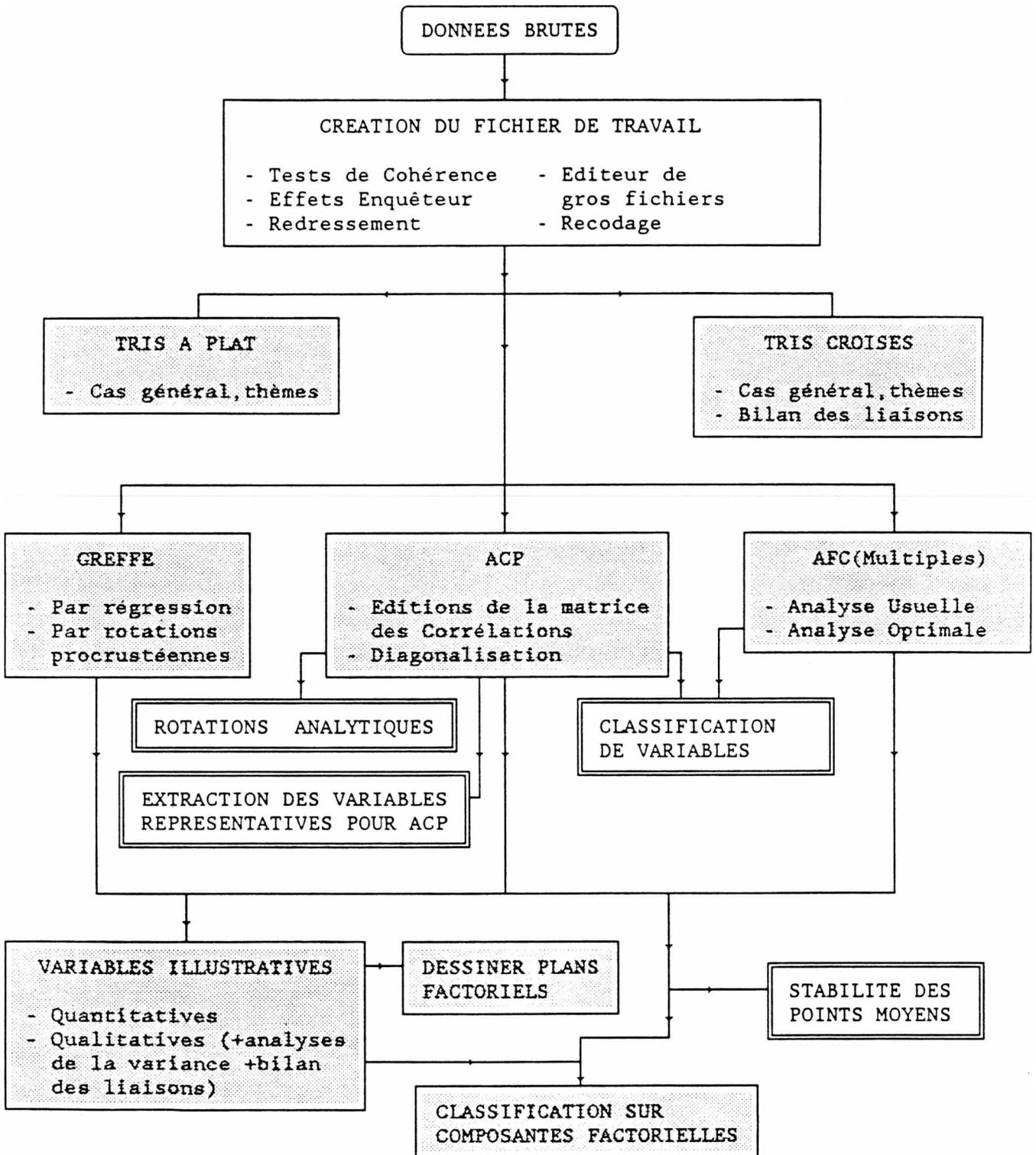


FIGURE 4.1 : Principaux modules du logiciel

#### 4.1. Création du fichier de travail

La première opération dans la chaîne des traitements est la création du fichier de travail à partir du fichier des données brutes (fichier dans son état de remise par l'institut du sondage). Il s'agit de corriger les erreurs de codage, de traiter des données manquantes, d'examiner la qualité des questionnaires et de l'échantillon, de générer, finalement, un fichier normalisé.

Un éditeur de gros tableaux de données combiné à une routine générale de recodage favorisant le traitement des variables par lot assure la partie "automatisée" du travail. Divers utilitaires permettent un premier examen de la qualité des questionnaires. Par exemple, décompte du nombre des "non-réponse" ; tests de cohérence. Remarquons que l'on applique ici des règles de "bon sens", issues de l'expérience dans le domaine, donc subjectives et peu généralisables.

Dans le cas des grosses enquêtes, l'examen détaillé de tous les questionnaires est un travail lourd et non nécessairement utile vu l'influence réduite qu'a individu isolé sur les résultats des analyses statistiques. Aussi, ayant vérifié dans la pratique l'importance de l'effet "enquêteur", est-ce à l'examen des questionnaires regroupés par enquêteur que nous procédons, dans un premier temps.

Quand on dispose des variables quantitatives, on a recours, pour caractériser globalement l'enquêteur  $E_\ell$ , à la distance Mahalanobis (métrique  $V^{-1}$ ). On mesure la proximité entre le centre de gravité  $\bar{x}_\ell$ , correspondant aux réponses obtenues par l'enquêteur  $E_\ell$ , et le centre de gravité global  $\bar{x}$ , en pondérant par l'effectif  $n_\ell$  :

$$D_\ell = n_\ell (\bar{x}_\ell - \bar{x})' V^{-1} (\bar{x}_\ell - \bar{x}) \quad (4.1)$$

où  $V$  est la matrice de variance calculée à partir de tous les  $n$  individus.

Le calcul fournit un classement des enquêteurs par ordre décroissant de l'indice  $D_\ell$ . Des valeurs élevées de  $D_\ell$  indiquent la présence de questionnaires qui sortent globalement du lot, sur lesquels il faut s'appesantir pour un examen approfondi ; éventuellement, on ira jusqu'à

eliminer le bloc entier. Un utilitaire permettant l'étude visuelle des proximités "intra-enquêteur", facilite ce travail. On constate, en effet, que certains enquêteurs ont le talent fâcheux de sélectionner des interviewés qui répondent aux questions presque à l'identique!

Soulignons ici, que la suppression des individus et, en particulier, des blocs d'individus provenant du même enquêteur, peut conduire à un échantillon déséquilibré qui ne respecte pas le plan initial de sondage (par exemple les stratifications par région ou par taille d'agglomération). Si tel est le cas, il faut recourir à des procédures de redressement de l'échantillon ; nous en proposons une qui s'appuie sur l'algorithme RAS [Thionet 1976].

Pour les variables qui jouent le rôle de variables actives en ACP, un traitement spécial est prévu pour les données manquantes (modalités "ne sait pas" ou "ne veut pas répondre"). Ces variables, que nous appellerons par la suite "thèmes", sont, dans les enquêtes Agoramétrie, des variables ordinales, codées en valeurs entières consécutives, dont les liaisons sont mesurées à l'aide du coefficient de corrélation. Ne pouvant maintenir dans l'ACP la modalité "non-réponse", nous l'éliminons de la façon suivante : pour chaque variable la moyenne  $m$  est calculée sur l'ensemble des réponses effectives ; les non-réponses sont alors affectées au hasard aux deux valeurs entières entourant  $m$ , proportionnellement à la distance de  $m$  à ces valeurs. De cette façon les individus recodés n'affectent guère les moyennes et exercent une influence quasiment nulle dans le calcul des corrélations.

#### Justification de la formule 4.1

Le  $n$ -échantillon (ensemble  $I$ ), réunion des  $m$   $n_\ell$ -sous-échantillons (ensembles  $I_\ell$ ), se trouve partitionné pour chaque enquêteur  $E_\ell$  en :  $I = I_\ell \cup I_{-\ell}$ , où  $I_{-\ell} = \bigcup_{i \neq \ell} I_i$ .

Notons  $\bar{x}_\ell$ ,  $\bar{x}_{-\ell}$  et  $V_\ell$ ,  $V_{-\ell}$  respectivement les  $p$ -vecteurs moyens et les matrices de variance correspondantes,  $W_\ell = [n_\ell V_\ell + (n - n_\ell) V_{-\ell}] / (n - 2)$  étant la matrice de variance "intra". La distance entre  $\bar{x}_\ell$  et  $\bar{x}_{-\ell}$  peut être mesurée à l'aide de l'indice (distance de Mahalanobis avec la métrique  $W_\ell^{-1}$ ) :

$$T_\ell = \frac{n_\ell(n - n_\ell)}{n} (\bar{x}_\ell - \bar{x}_{-\ell})' W_\ell^{-1} (\bar{x}_\ell - \bar{x}_{-\ell})$$

Sous les hypothèses de la normalité des observations et de l'égalité des matrices de variance théoriques  $\Sigma_\ell$  et  $\Sigma_{-\ell}$ , la statistique  $T_\ell$  permet de tester l'égalité des moyennes théoriques  $\bar{\mu}_\ell$  et  $\bar{\mu}_{-\ell}$  dont  $\bar{x}_\ell$  et  $\bar{x}_{-\ell}$  sont des estimations ; elle est distribuée [Anderson 1984, p.167] comme un  $T^2$  de Hotelling à  $(p, n-2)$  degrés de liberté. En calculant pour chaque groupe d'interviewés (un par enquêteur) la statistique  $T_\ell$ , on aurait donc pu espérer repérer les lots d'interviews "hors du commun". Mais on a préféré procéder autrement.

Il faut savoir, en effet, que les enquêteurs ne peuvent enquêter qui ils veulent ; ils doivent respecter leurs "feuilles de route" qui, imposant de sélectionner les interviewés selon des contraintes draconiennes, sont différentes pour chacun. Aussi un groupe d'interviewés ne saurait être assimilé à un "échantillon", et cela quelle que soit l'hypothèse qui pourrait être faite au niveau de la loi parente... Et il est normal, par ailleurs, que l'on note des différences entre les moyennes  $\bar{x}_\ell$  associées aux différents groupes. Chaque enquêteur participe à la reconstruction, comme dans un "puzzle", d'une même loi parente. Il n'y a pas q "échantillons" ; il n'y a qu'un seul : celui formé par les n interviewés.

Assimilons alors l'échantillon amputé de sa partie  $I_\ell$ , qui est suffisamment grand et dans tous les cas beaucoup plus grand que  $I_\ell$ , à une population de Laplace-Gauss  $N(\bar{x}_{-\ell}; V_{-\ell})$ . Pour juger du caractère "hors du commun" des interviewés du  $\ell^{\text{ème}}$  groupe, il est logique, alors, d'avoir recours à l'indice (métrique  $V_{-\ell}^{-1}$ )

$$\Delta_\ell = n_\ell (\bar{x}_\ell - \bar{x}_{-\ell})' V_{-\ell}^{-1} (\bar{x}_\ell - \bar{x}_{-\ell}) \quad (4.2)$$

qui se comporte comme un khi-2 à p degrés de liberté. Mais on a préféré, pour éviter des complications qui se révélèrent inutiles puisque nous sélectionnons les extrêmes en procédant par comparaisons, avoir recours à l'indice  $D_\ell$  de la formule 4.1 (métrique  $V^{-1}$ ). Il faut ajouter que, quand  $n_\ell$  est nettement inférieur à n pour tout  $\ell$ , les trois indices  $D_\ell$ ,  $T_\ell$  et  $\Delta_\ell$  conduisent pratiquement aux mêmes résultats.

La démarche qui a conduit à l'indice  $D_\ell$  pour repérer des groupes d'individus bizarres illustre bien le "pragmatisme" qui est à la base des propositions que l'on trouve dans cette thèse.

#### 4.2 Tris à plat

Le module (à un seul passage) de tris à plat fournit des documents comme celui de la figure 4.2.a dans le cas général, ou comme celui de la figure 4.2.b quand il s'agit de variables-thèmes. On voit apparaître ici le problème de gestion que posent les libellés des questions et des modalités des réponse.

#### 4.3. Tris croisés

Lors du traitement d'une enquête faisant intervenir plusieurs centaines de variables qualitatives, le module des tris croisés est, de loin, le plus utilisé. La bonne qualité des sorties apparaît ici, plus qu'ailleurs, essentielle. Aussi, dans ce module, nous proposons une diversité de présentations soignées et bien adaptées à la nature des variables traitées. Le cas général où la variable "en colonnes" contient un nombre élevé de modalités (relativement à la largeur du papier) est illustré dans la figure 4.3. Quand la variable en colonnes est du type "thème", on fait apparaître (figure 1.1), en fin de chaque ligne du tableau, la moyenne de cette variable entière calculée pour ces individus prenant la modalité-ligne correspondante ; cela permet une caractérisation rapide, par comparaisons, de la liaison entre la variable-thème et la variable qualitative (la tendance politique, par exemple, dont les modalités apparaissent en lignes).

L'intensité de la liaison entre deux variables est mesurée à l'aide du test du khi-2 effectué sur le tableau de contingence. La probabilité de dépassement, qui est ici calculée avec une grande précision, n'est pas utilisée pour rejeter l'hypothèse d'indépendance mais, conformément à nos pratiques, seulement comme un indicateur de la significativité des liaisons.

Avec la procédure à laquelle on a recours dans le module des tris croisés, on est capable de construire rapidement plusieurs milliers de tableaux croisés en un seul passage. Sans imprimer, de préférence, ces tableaux, on travaille, dans un premier temps, seulement sur les indi-

LIBELLE	EFFEC	%
<b>Politique</b> " En politique, vous situez-vous plutôt à .... "		
EXTREME GAUCHE	14	1.0
GAUCHE	472	32.1
CENTRE	254	17.3
DROITE	314	21.4
EXTREME DROITE	14	1.0
INDIFFERENT	209	14.2
NON REPONSE	193	13.1
<b>Proximité politique</b> " De quel parti politique vous sentez-vous le plus proche ? "		
EXTREME GAUCHE	11	0.7
P.C.F.	62	4.2
P.SOCIALISTE	430	29.3
RADICAUX GAUCHE	31	2.1
ECOLOGISTES	66	4.5
U.D.F.	165	11.2
R.P.R.	196	13.3
FRONT NATIONAL	41	2.8
AUCUN	320	21.8
NON REPONSE	148	10.1
<b>Personnalité politique préférée</b> " Parmi les personnalités politiques suivantes, laquelle préférez-vous ? "		
R.BARRE	194	13.2
J.CHABAN DELMAS	29	2.0
J.P.CHEVENEMENT	11	0.7
J.CHIRAC	121	8.2
J.DELORS	40	2.7
L.FABIUS	61	4.1
V.GISCARD D'ESTAING	66	4.5
L.JOSPIN	28	1.9
B.LALONDE	24	1.6
F.LEOTARD	84	5.7
J.M.LE PEN	43	2.9
G.MARCHAIS	33	2.2
F.MITERRAND	272	18.5
M.ROCARD	153	10.4
S.VEIL	38	2.6
AUCUNE	149	10.1
REFUS DE REpondre	124	8.4

figure 4.2.a

LIBELLE	EFF.	%	Pour/ Contre
<b>THEME N° 70 ACCIDENTS DU TRAVAIL RARES</b>			
PAS DU TOUT D'ACCORD	432	29.4	74.8
PAS TELLEMENT D'ACCORD	668	45.4	
PEUT-ETRE D'ACCORD	182	12.4	
BIEN D'ACCORD	151	10.3	
ENTIEREMENT D'ACCORD	37	2.5	12.8
<b>THEME N° 77 POUR LES 35 HEURES</b>			
PAS DU TOUT D'ACCORD	322	21.9	38.2
PAS TELLEMENT D'ACCORD	239	16.3	
PEUT-ETRE D'ACCORD	322	21.9	
BIEN D'ACCORD	236	16.1	
ENTIEREMENT D'ACCORD	351	23.9	39.9
<b>THEME N° 79 SE PASSER DE VOITURE</b>			
PAS DU TOUT D'ACCORD	518	35.2	62.1
PAS TELLEMENT D'ACCORD	395	26.9	
PEUT-ETRE D'ACCORD	284	19.3	
BIEN D'ACCORD	157	10.7	
ENTIEREMENT D'ACCORD	116	7.9	18.6
<b>THEME N° 83 MONTEE DE LA VIOLENCE</b>			
PAS DU TOUT D'ACCORD	80	5.4	16.5
PAS TELLEMENT D'ACCORD	162	11.0	
PEUT-ETRE D'ACCORD	169	11.5	
BIEN D'ACCORD	429	29.2	
ENTIEREMENT D'ACCORD	630	42.9	72.0
<b>THEME N° 84 SYNDICATS INDISPENSABLES</b>			
PAS DU TOUT D'ACCORD	171	11.6	27.3
PAS TELLEMENT D'ACCORD	231	15.7	
PEUT-ETRE D'ACCORD	322	21.9	
BIEN D'ACCORD	376	25.6	
ENTIEREMENT D'ACCORD	370	25.2	50.7
<b>THEME N° 88 EMPLOUER DANS LES SERVICES PUBLICS</b>			
PAS DU TOUT D'ACCORD	196	13.3	27.7
PAS TELLEMENT D'ACCORD	211	14.4	
PEUT-ETRE D'ACCORD	302	20.5	
BIEN D'ACCORD	389	26.5	
ENTIEREMENT D'ACCORD	372	25.3	51.8
<b>THEME N° 89 POUR L'INTERVENTION POLITIQUE DU PAPE</b>			
PAS DU TOUT D'ACCORD	661	45.0	66.9
PAS TELLEMENT D'ACCORD	322	21.9	
PEUT-ETRE D'ACCORD	216	14.7	
BIEN D'ACCORD	146	9.9	
ENTIEREMENT D'ACCORD	125	8.5	18.4
<b>THEME N° 90 ON N'APPREND PLUS RIEN A L'ECOLE</b>			
PAS DU TOUT D'ACCORD	389	26.5	57.3
PAS TELLEMENT D'ACCORD	454	30.9	
PEUT-ETRE D'ACCORD	270	18.4	
BIEN D'ACCORD	205	13.9	
ENTIEREMENT D'ACCORD	152	10.3	24.3

figure 4.2.b

Enquête Agoramétrie 1987, n=1470 individus

FIGURE 4.2 : Tris à plat

Politique								
Abréviations utilisées et tri à plat :								
ABR	LIBELLE			EFFEC	%			
EX.G	EXTREME GAUCHE			14	1.0			
GAUC	GAUCHE			472	32.1			
CENT	CENTRE			254	17.3			
DROI	DROITE			314	21.4			
EX.D	EXTREME DROITE			14	1.0			
INDI	INDIFFERENT			209	14.2			
N.R.	NON REPONSE			193	13.1			

Religion	EX.G	GAUC	CENT	DROI	EX.D	INDI	N.R.	TOTAL
E F F E C T I F S								
SANS RELIGION	9	138	19	27	2	35	28	258
CATHO.NON PRATIQUANT	1	215	114	112	9	96	79	626
PRAT.OCCASIONNEL	0	63	73	102	3	43	49	333
PRAT.REGULIER	1	19	36	70	0	12	16	154
AUTRES CAS	3	37	12	3	0	23	21	99
TOTAL	14	472	254	314	14	209	193	1470
chi-2 - 240.18 ddl - 24 Proba - 1.4E-037								
P R O F I L S								
SANS RELIGION	35	535	74	105	8	136	109	1000
CATHO.NON PRATIQUANT	2	343	182	179	14	153	126	1000
PRAT.OCCASIONNEL	0	189	219	306	9	129	147	1000
PRAT.REGULIER	6	123	234	455	0	78	104	1000
AUTRES CAS	30	374	121	30	0	232	212	1000

Revenu	EX.G	GAUC	CENT	DROI	EX.D	INDI	N.R.	TOTAL
E F F E C T I F S								
- DE 4000F.	1	44	23	29	2	29	18	146
4000F. A -5000F.	1	39	18	23	1	27	23	132
5000F. A -6000F.	5	66	18	25	1	20	16	151
6000F. A -7000F.	1	59	30	25	2	28	19	164
7000F. A -8000F.	2	52	22	28	1	21	18	144
8000F. A -10000F.	1	83	44	38	2	31	18	217
10000F.A -15000F.	1	67	42	63	2	20	24	219
+ 15000F.	1	36	31	37	2	9	10	126
NON REPONSE	1	26	26	46	1	24	47	171
TOTAL	14	472	254	314	14	209	193	1470
chi-2 - 121.74 ddl - 48 Proba - 2.5E-008								
P R O F I L S								
- DE 4000F.	7	301	158	199	14	199	123	1000
4000F. A -5000F.	8	295	136	174	8	205	174	1000
5000F. A -6000F.	33	437	119	166	7	132	106	1000
6000F. A -7000F.	6	360	183	152	12	171	116	1000
7000F. A -8000F.	14	361	153	194	7	146	125	1000
8000F. A -10000F.	5	382	203	175	9	143	83	1000
10000F.A -15000F.	5	306	192	288	9	91	110	1000
+ 15000F.	8	286	246	294	16	71	79	1000
NON REPONSE	6	152	152	269	6	140	275	1000

Enquête Agoramétrie 1987, n=1470 individus

FIGURE 4.3 : Extrait de tris croisés

cateurs que sont les probabilités de dépassement (sauvegardées dans un fichier). En consultant des tableaux-résumé comme celui de la figure 4.4, on peut alors dans un deuxième temps, sans se perdre dans la masse, circuler de façon optimale parmi les tris croisés pour se concentrer sur les faits les plus probants. Il faut noter que, dans la figure 4.4, le choix des seuils de significativité, qui interpelle toujours le statisticien, repose entièrement sur l'expérience acquise.

C'est de cette façon que nous arrivons à effectuer, dans des très courts délais, la première analyse d'un ensemble considérable de tableaux : dans nos enquêtes, il est issu des croisements entre les quelque deux à trois cents variables qualitatives. Un environnement conversationnel adéquat (choix des variables à croiser et des différentes options à l'aide d'une souris), facilite grandement le choix.

#### Remarque sur le calcul du khi-2 et le choix des seuils

Ayant traité de nombreuses grosses enquêtes, nous avons compris l'intérêt qu'il y a à calculer avec précision l'ordre de grandeur de la valeur  $P$  de la probabilité de dépassement. En effet, pour des échantillons de grande taille (1500 ou plus), on peut théoriquement obtenir des valeurs de  $P$  de l'ordre de  $10^{-100}$ , ou inférieures, caractérisant une très forte liaison. Une valeur de l'ordre de  $10^{-3}$  n'est, ici, guère significative. Certes, la distribution de la statistique considérée n'est qu'asymptotiquement un khi-2 ; et l'on peut s'interroger sur la significativité réelle des valeurs de  $P$  quand on se situe en "queue de distribution"... Mais rappelons que ces valeurs sont assimilées à des indices permettant de classer les liaisons en différentes catégories, des plus fortes jusqu'aux plus faibles. L'essentiel, ici, est de respecter avec  $P$  un ordre immuable suivant lequel on décide (en faisant à tout coup des erreurs, mais en nombre minimum) de ranger les liaisons : la fonction fournissant  $P$  (la distribution du khi-2) doit être croissante avec l'écart au sens du khi-2 à l'indépendance, et, fournir un résultat ( $P$ ) qui donne, au moins, une idée à l'analyste de la force de la liaison.

Cette règle pragmatique de rangement des tableaux croisés en catégories, autant l'appliquer avec rigueur. Et ceci nécessite, compte tenu de la taille des échantillons considérés, le calcul précis de très

THEMES	REG.	HAB.	SEXE	AGES	REL.	DIP.	CSP	POL.	PRPO	TOTAL
LIBERALISER L'AVORTEMENT	■	■		■	■	■	■	■	■	8
PUBLICITE INDISPENSABLE			■							1
CONSTRUIRE DES CENTRALES NUCLEAIRES			■					■	■	3
CONFIANCE AUX MEDECINS				■	■	■		■	■	4
DIEU EXISTE	■		■	■	■	■	■	■	■	8
EGALISER LES REVENUS	■					■	■	■	■	5
CONFIANCE EN LA JUSTICE				■	■		■			3
VIVRE COMME AUX USA	■			■	■			■		4
POUR LA FORCE DE FRAPPE			■	■	■			■	■	5
ETUDIANTS PARASITES		■				■	■	■	■	5
TROP DE TRAVAILLEURS IMMIGRES	■	■		■	■	■	■	■	■	8
LIMITER LES HERITAGES	■							■	■	3
CONTRE LA PORNOGRAPHIE		■	■	■	■	■	■	■	■	8
PRIS POUR ABRUTIS A LA TELEVISION						■		■	■	0
AIDER LES PAYS SOUS-DEVELOPPES	■				■	■			■	5
POLLUTION PREOCCUPANTE	■								■	2
POUR LE PROGRES TECHNIQUE										0
POUR LA PATRIE		■		■	■	■	■	■	■	7
ENCOURAGER LA NATALITE					■		■	■	■	4
PROGRES SCIENTIFIQUE NEFASTE						■			■	1
POUR LA FAMILLE	■	■		■	■	■	■	■	■	8
CONTRE LE TRAVAIL	■	■		■	■		■	■	■	7
DEFENSE DU CONSOMMATEUR										0
CONTRE LE MARIAGE				■	■	■	■	■	■	6
MAINTENIR LA CROISSANCE					■	■		■	■	4
REDUIRE LES DEPENSES MILITAIRES				■	■		■	■	■	5
CENSURER CERTAINS LIVRES	■	■		■	■	■	■	■	■	8
GOVERNEMENT INEFFICACE					■			■	■	3
RESPECTER LES CONVENANCES		■		■	■	■	■	■	■	7
SENTIMENT D'INSECURITE	■	■	■	■	■	■	■	■	■	9
NE PAS HESITER A S'ENDETTER					■		■			3
CRISE DE L'ENERGIE PREOCCUPANTE			■			■				2
LIMITER LE DEVELOPPEMENT DE L'INDUSTRIE	■					■				2
UTILISER L'ENERGIE SOLAIRE										0
SOUTENIR LES ECOLOGISTES		■			■			■	■	4
PAS DE PERMISSION DE SORTIE	■	■		■	■	■	■	■	■	8
ACCIDENTS DU TRAVAIL RARES								■	■	2
POUR LES 35 HEURES	■	■		■	■		■	■	■	7
SE PASSER DE VOITURE	■	■								2
MONTEE DE LA VIOLENCE	■	■		■	■	■	■	■	■	8
SYNDICATS INDISPENSABLES					■			■	■	3
EMBAUCHER DANS LES SERVICES PUBLICS						■	■	■	■	4
POUR L'INTERVENTION POLITIQUE DU PAPE					■			■		2
ON N'APPREND PLUS RIEN A L'ECOLE						■	■	■		3
AGRICULTEURS TROP SUBVENTIONNES	■	■					■			3
ORDINATEURS MENACENT NOS LIBERTES			■			■	■	■	■	4
MEDECINE GRATUITE		■		■	■		■	■	■	6
GUERRE MONDIALE INEVITABLE			■			■	■			3
RETRAITE PLUS JEUNE				■	■	■	■	■	■	6
FRAUDER LE FISC					■	■			■	3
TROP DE GREVES		■		■	■	■	■	■	■	7
HOMOSEXUELS COMME LES AUTRES		■		■	■	■	■	■	■	7
TROP DE FONCTIONNAIRES								■	■	2
HOMMES POLITIQUES INTEGRES				■	■	■	■	■	■	6
MARCHE COMMUN FAVORABLE A LA FRANCE							■			0
INDUSTRIE CHIMIQUE POLLUANTE							■			1
RETABLIR LA PEINE DE MORT	■	■		■	■	■	■	■	■	8
EXAGERATION DES DANGERS DE L'ALCOOL			■							1
NECESSITE D'EXPERIMENTER SUR ANIMAUX	■	■	■							3
HASCHISCH EN VENTE LIBRE	■	■		■	■	■	■	■	■	8
CONTROLES D'IDENTITE INDISPENSABLES	■	■		■	■	■	■	■	■	8

FIGURE 4.4 : Bilan de tris croisés (Suite...)

THEMES	REG.	HAB.	SEXE	AGES	REL.	DIP.	CSP	POL.	PRPO	TOTAL
LIMITER LES IMPORTATIONS JAPONAISES				■	■	■	■			4
ACHETER DU GAZ AUX SOVIETIQUES			■					■	■	3
REDUIRE LES INDEMNITES DU CHOMAGE	■				■			■	■	4
SOUTENIR LES PACIFISTES		■			■			■	■	4
L'EUROPE NE MARCHERA JAMAIS						■	■		■	3
POUR DES TELES PRIVEES		■	■			■		■	■	5
NON A CEUX QUI BLOQUENT LE PAYS	■	■		■	■	■	■	■	■	8
NE PAS COUVRIR LES PETITES MALADIES								■	■	2
DEBAT POLITIQUE DEPASSE										0
TROP DE CHARGES SUR LES ENTREPRISES		■			■		■	■	■	5
POUR L'AUTO-DEFENSE		■			■	■	■	■		5
MOINS D'ETAT			■					■	■	3
IL FALLAIT DES CENTRALES NUCLEAIRES			■	■			■	■	■	5
ENTREPRISES NATIONALISEES EFFICACES			■					■	■	3
VOTE MUNICIPAL AUX ETRANGERS	■	■		■	■	■		■	■	7
DENATIONALISER RENAULT			■		■	■		■	■	5
INDEPENDANCE A LA NOUVELLE CALEDONIE	■				■			■	■	4
EXPORTER LES ARMES FRANCAISES		■	■	■	■			■	■	6
MOINS DE ROBOTS	■	■	■			■	■	■		6
POUR DES FEMMES NUES A LA TELE			■	■	■	■	■	■	■	7
LIBERTE DES PRIX			■				■	■	■	4
MITTERRAND ET CHIRAC, CA MARCHE BIEN		■						■	■	3
POUR SUIVRE LES CRIMINELS NAZIS									■	1
GORBATCHEV INSPIRE CONFIANCE								■	■	2
LIBERTE DES LOYERS		■	■	■	■		■	■	■	7
VERITE CACHEE EN FRANCE SUR TCHERNOBYL										0
JAMAIS NEGOCIER AVEC LES TERRORISTES				■				■	■	3
ISOLER MALADES DU SIDA	■	■		■	■	■	■	■	■	8
LA POLICE REMPLIT SA MISSION				■	■	■	■	■	■	6
AIDER IRAK CONTRE IRAN			■						■	2
DETTE DU TIERS-MONDE PREOCCUPANTE					■		■	■	■	0
TRAVAILLER COMME AU JAPON		■		■	■			■	■	6
LA PRIVATISATION, C'EST BIEN			■		■			■	■	4
PAS SE PLAINDRE SI TRAVAIL	■	■		■	■	■	■	■	■	8
ARMEE FRANCAISE MEILLEURE				■	■	■	■	■	■	6
PRISON POUR ALCOOL AU VOLANT			■	■	■		■			4
POUR L'IMMIGRATION	■	■			■	■		■	■	6
CONFIANCE AUX PSYCHIATRES				■						1
POUR L'ASTROLOGIE										0
PAYER LES CHEQUES										0
TRANSPORTS DANGEREUX MAL CONTROLES										0
CONTRE DISNEYLAND		■		■			■			3
TOTAL	30	38	25	44	56	46	51	71	69	430

Significativité des liaisons :

■ : à  $10^{-3}$ ■ ■ : à  $10^{-4}$ ■ ■ ■ : à  $10^{-6}$ ■ ■ ■ ■ : à  $10^{-12}$ ■ ■ ■ ■ ■ : à  $10^{-30}$ 

LISTE DES ABREVIATIONS :

REG. : REGION  
DIP. : DIPLOMEHAB. : HABITAT  
POL. : TENDANCE POLITIQUEREL. : RELIGION  
PRPO : PROXIMITE POLITIQUE

Enquête Agoramétrie 1987, n=1470 individus

FIGURE 4.4 : (...Suite)

faibles valeurs de  $P$  ( jusqu'à  $10^{-308}$ , limite des réels en double précision codés IEEE).

Aussi nous procédons de la façon suivante : soit  $\chi^2$  la valeur prise par la statistique du khi-2 à  $\nu$  degrés de liberté et soit  $P(\chi^2|\nu)$  la valeur de la probabilité de dépassement correspondante. Pour  $\nu$  pair, la valeur exacte de  $P(\chi^2|\nu)$  est calculée comme la somme d'un petit nombre ( $(\nu)/2$ ) de termes positifs (voir formule 26.4.5 dans [Zelen et al. 1964]). Dans ce cas donc, on a la précision souhaitée en rapport avec l'ordre de grandeur de  $P$ . Pour  $\nu$  impair, avec le calcul direct, la précision n'est correcte qu'au deçà de  $10^{-8}$  (formules 26.4.4 et 26.2.17 dans [Zelen et al. 1964]). Aussi, quand on trouve, en passant par la formule normale, pour un nombre de degrés de liberté impair que  $P(\chi^2|2\nu+1)$  est inférieur à  $10^{-7}$ , on a recours à une interpolation logarithmique pour obtenir la valeur de la probabilité :

$$\log[P(\chi^2|2\nu+1)] = (\log[P(\chi^2|2\nu)] + \log[P(\chi^2|2\nu+2)]) / 2.$$

#### 4.4. Analyse en Composantes Principales

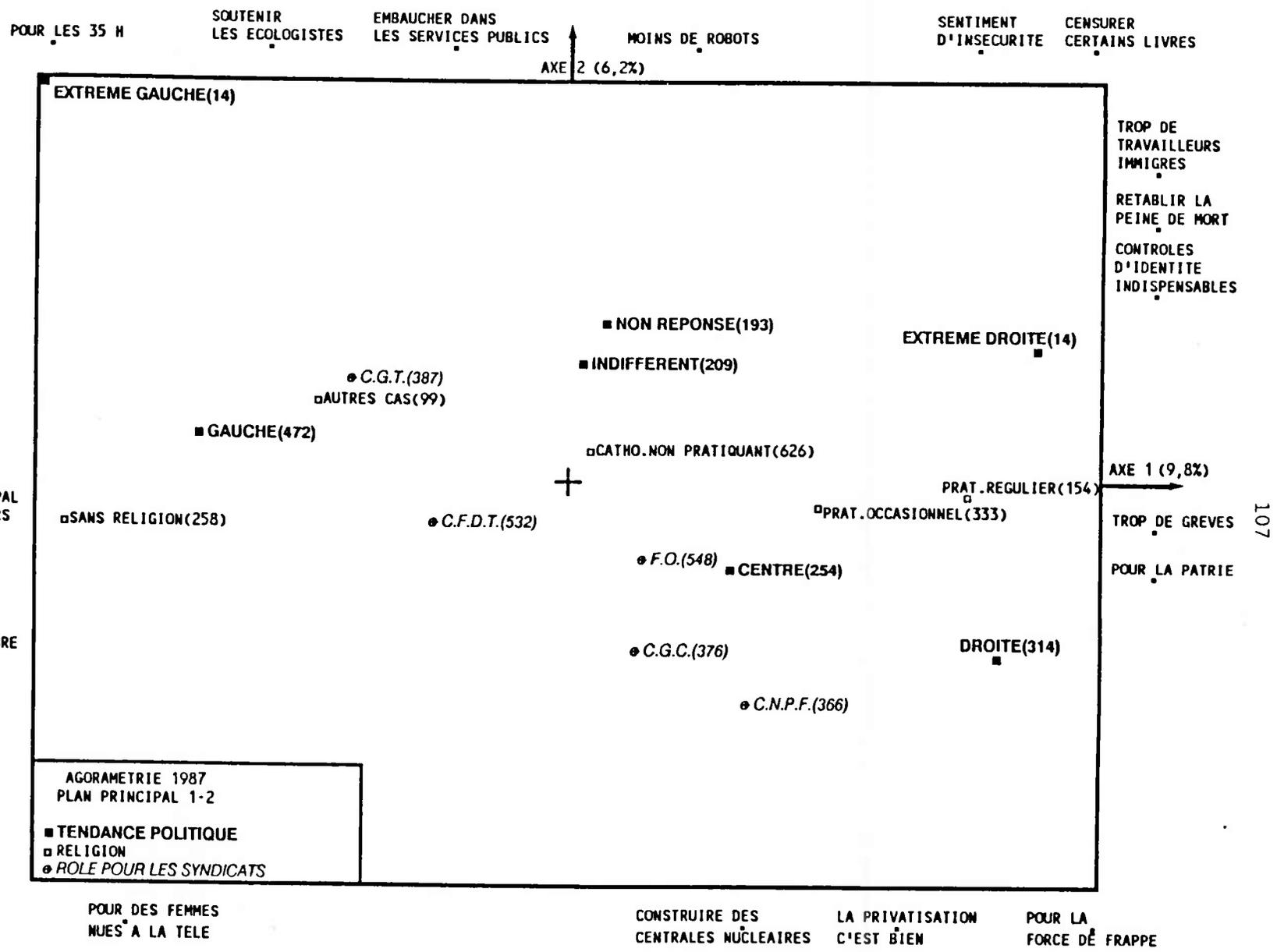
La procédure de l'ACP s'appuie sur un algorithme classique de diagonalisation qui consiste à appliquer la méthode Q-R sur la réduction tridiagonale (obtenue par la méthode de Householder) d'une matrice symétrique [Wilkinson 1965]<sup>1</sup>. Au niveau des documents statistiques l'accent a été mis sur la présentation des plans principaux et des cercles de corrélation. Un exemple de plan principal est donné dans la figure 4.5 où, pour orienter le plan  $(u_1, u_2)$ , on y a projeté les vecteurs unitaires associés à des variables caractéristiques des composantes principales correspondantes  $c^1, c^2$ .

---

<sup>1</sup>Pour des références plus récentes sur les algorithmes de diagonalisation, on pourra consulter [Golub 1985]. Notons toutefois que notre (relativement) ancien algorithme de type Q-R (qui implique donc le calcul obligatoire de la totalité des valeurs et vecteurs propres) a des performances tout à fait satisfaisantes : programmé en langage C sur un micro-ordinateur compatible IBM-PC haut de gamme, il demande moins de 1" pour diagonaliser une matrice (20x20), 10" pour une matrice (50x50) et 1'20" pour une matrice (100x100).

Enquête Agoramétrie 1987, n=1470 individus

FIGURE 4.5 : Plan principal



Un module conversationnel séparé permet, à partir des coordonnées mises sur fichier, le positionnement des libellés sur les graphiques à l'endroit souhaité autour du point de repère respectant les coordonnées exactes. Ayant constaté que la façon de disposer les libellés sur un graphique dépend de considérations subjectives, et que l'interprétation des indicateurs factoriels s'en trouve parfois affectée, nous n'avons pas cherché à programmer une procédure automatique de positionnement ; d'autant que la mise au point d'un algorithme efficace pour la gestion des recouvrements (cas où des libellés se superposent) ne semble guère évidente.

Deux modules importants, concernant la sélection des variables actives et le nombre d'axes à retenir, viennent respectivement en amont et en aval du module principal de l'ACP dans la chaîne des traitements. Le premier fournit une présentation intéressante de la matrice des corrélations ; le second permet un examen rapide des liaisons globales entre variables du signalétique et composantes factorielles.

#### 4.4.1. Présentations de la matrice des corrélations - Choix des variables actives.

Une présentation classique de la matrice des corrélations est celle de la figure 4.6, où on aurait pu faire apparaître seulement la partie triangulaire. Pour faciliter la lecture, le programme encadre automatiquement les corrélations supérieures à 0.3 en valeur absolue. Lorsque la dimension de la matrice est grande (nous manipulons souvent une centaine de variables), cette présentation se révèle insuffisante. On préfère alors, à la place, celle de la figure 4.7 où pour chaque variable apparaissent :

- les variables (10 ici) qui lui sont le plus corrélées,
- la moyenne,
- l'écart-type,
- le coefficient de corrélation multiple  $R$  (noté c.c.m.), qui traduit la liaison que la variable entretient avec l'ensemble des autres variables
- l'histogramme, lorsqu'il s'agit des variables-"thèmes".

	1	2	4	5	6	7	9	10	11
	AVORT	PUBLICIT	NUCLEAIR	MEDECINS	DIEU	REVENUS	JUSTICE	USA	FRAPPE
1 AVORT	1.0	0.131	0.047	-0.068	-0.341	0.042	-0.041	0.008	-0.029
2 PUBLICIT	0.131	1.0	0.191	0.104	-0.008	-0.160	0.137	0.091	0.172
4 NUCLEAIR	0.047	0.191	1.0	0.108	0.035	-0.135	0.144	0.093	0.284
5 MEDECINS	-0.068	0.104	0.108	1.0	0.176	0.045	0.321	0.099	0.122
6 DIEU	-0.341	-0.008	0.035	0.176	1.0	-0.010	0.103	0.075	0.113
7 REVENUS	0.042	-0.160	-0.135	0.045	-0.010	1.0	0.034	-0.008	-0.132
9 JUSTICE	-0.041	0.137	0.144	0.321	0.103	0.034	1.0	0.110	0.136
10 USA	0.008	0.091	0.093	0.099	0.075	-0.008	0.110	1.0	0.145
11 FRAPPE	-0.029	0.172	0.284	0.122	0.113	-0.132	0.136	0.145	1.0
13 ETUDIANT	-0.120	-0.067	0.018	-0.038	0.090	0.007	-0.073	0.182	0.044
15 IMMIGRES	-0.166	-0.011	0.085	0.087	0.196	-0.028	-0.059	0.207	0.236
16 HERITAGE	-0.011	-0.050	-0.077	-0.040	-0.036	0.219	0.052	0.099	-0.083
18 PORNO.	-0.374	-0.114	-0.044	0.125	0.359	0.112	0.116	0.022	0.076
19 TELEVI.	-0.084	-0.130	-0.074	-0.151	-0.011	0.046	-0.180	0.016	-0.006
21 SOUS-DEV	0.008	-0.040	-0.052	0.036	0.035	0.121	0.137	-0.042	-0.129
22 POLLUT.	0.001	-0.034	-0.159	0.019	0.025	0.159	-0.083	-0.083	-0.067
23 PROG TEC	0.084	0.150	0.122	0.091	-0.013	-0.004	0.150	0.254	0.058
26 PATRIE	-0.247	0.021	0.126	0.220	0.293	-0.008	0.239	0.164	0.250
28 NATALITE	-0.167	0.101	0.111	0.165	0.169	-0.063	0.212	0.146	0.177
29 SCIENTIF	-0.106	-0.111	-0.176	-0.036	0.078	0.086	-0.059	0.080	-0.073
30 FAMILLE	-0.243	0.029	0.095	0.219	0.301	-0.000	0.107	0.059	0.215
31 TRAVAIL	0.215	-0.002	-0.092	-0.128	-0.163	0.129	-0.090	-0.003	-0.173
33 CONSOM.	0.033	-0.017	-0.097	0.052	0.036	0.145	0.055	-0.015	0.017
34 MARIAGE	0.260	0.007	-0.025	-0.174	-0.249	0.045	-0.113	0.070	-0.065
35 CROISS.	-0.025	0.157	0.143	0.179	0.108	-0.039	0.131	0.119	0.281
36 DEP MIL.	0.138	-0.052	-0.246	-0.086	-0.131	0.169	-0.104	-0.081	-0.339
41 CENSURE	-0.355	-0.098	-0.053	0.134	0.307	0.102	0.096	0.058	0.069
42 GOUV. IN.	0.081	-0.106	-0.154	-0.064	-0.092	0.182	-0.185	0.008	-0.065
43 CONVEN.	-0.165	0.036	0.109	0.174	0.246	0.021	0.103	0.101	0.228
46 INSECUR.	-0.182	-0.024	0.004	0.129	0.195	0.050	-0.052	0.122	0.095
56 ENDETTER	0.175	0.102	0.087	-0.049	-0.142	-0.060	-0.002	0.063	0.002
57 ENERGIE	-0.032	-0.021	-0.055	0.074	0.098	0.106	0.011	0.097	-0.008
59 INDUST.	-0.020	-0.065	-0.102	0.020	-0.006	0.113	-0.002	0.076	-0.087
60 SOLAIRE	0.102	0.024	-0.000	0.002	-0.016	0.098	0.008	-0.035	0.009
66 ECOLOGIE	0.094	-0.070	-0.242	-0.026	-0.066	0.207	-0.001	-0.063	-0.202
67 P. SORTIE	-0.191	-0.015	0.052	0.162	0.189	-0.012	0.005	0.117	0.175
70 ACC. TRAV	-0.047	0.003	0.126	0.060	0.047	-0.066	0.172	0.107	0.120
77 35H.	0.175	-0.035	-0.148	-0.057	-0.168	0.243	-0.024	-0.011	-0.179
79 VOITURE	0.011	-0.154	-0.138	-0.078	0.009	0.066	0.002	-0.078	-0.121
83 SYNDICAT	0.132	0.001	-0.034	0.105	-0.049	0.198	0.125	0.008	-0.053
88 EMBAUC. P	0.047	-0.071	-0.134	0.091	-0.049	0.245	0.007	0.062	-0.097
89 PAPE	-0.074	0.033	0.046	0.088	0.217	0.006	0.141	0.152	0.110
90 R. ECOLE	-0.090	-0.068	-0.013	-0.033	0.047	-0.027	-0.095	0.106	0.060
92 AGRICUL.	0.020	-0.065	0.013	-0.005	-0.044	0.060	-0.034	0.044	-0.018
93 ORDINAT.	-0.105	-0.167	-0.168	-0.044	0.095	0.152	-0.065	0.026	-0.049
95 MED. GRAT	0.110	-0.096	-0.165	-0.030	0.128	0.210	-0.087	0.008	-0.155
96 GUERRE	-0.116	-0.107	-0.134	0.027	0.140	0.108	-0.065	0.146	-0.058
99 RETRAITE	0.074	-0.022	-0.117	-0.058	-0.146	0.166	-0.071	0.012	-0.129
100 FISC	0.120	0.015	-0.017	-0.117	-0.101	-0.032	-0.118	0.034	-0.031
101 GREVES	-0.180	0.060	0.138	0.094	0.254	-0.177	0.025	0.131	0.214
103 HOMOS	0.305	0.090	-0.010	-0.030	-0.210	-0.005	-0.030	-0.077	-0.105
104 FONCTION	-0.083	0.091	0.118	-0.016	0.097	-0.135	-0.045	0.115	0.103
106 HOM. POL.	-0.016	0.088	0.109	0.271	0.084	0.006	0.352	0.123	0.139
109 C.E.E.	0.028	0.103	0.038	0.083	-0.015	-0.041	0.178	0.069	0.083
112 I. CHIMIE	0.014	-0.076	-0.213	-0.035	0.035	0.144	-0.105	-0.047	-0.074
117 PEI. MORT	-0.198	0.004	0.125	0.158	0.168	-0.059	-0.005	0.171	0.234
118 D. ALCOOL	0.047	-0.006	-0.006	0.026	-0.078	-0.008	-0.033	0.081	-0.015
121 ANIMAUX	-0.014	0.048	0.197	0.117	0.019	-0.051	0.087	0.015	0.174
122 HASCHISH	0.262	0.006	-0.026	-0.164	-0.204	0.020	-0.095	-0.034	-0.113
128 IDENTITE	-0.230	0.020	0.078	0.178	0.201	-0.029	0.096	0.116	0.235

Enquête Agoramétrie 1987, n=1470 individus.

FIGURE 4.6 : Extrait d'une matrice de corrélation (96x96)

**\*\* 15 : TROP DE TRAVAILLEURS IMMIGRES**

HISTOGRAMME		PRINCIPALES CORRELATIONS	
Modalité 1 :	218	RETABLIR LA PEINE DE MORT	0.530
Modalité 2 :	199	VOTE MUNICIPAL AUX ETRANGERS	-0.512
Modalité 3 :	308	PAS DE PERMISSION DE SORTIE	0.418
Modalité 4 :	261	SENTIMENT D'INSECURITE	0.414
Modalité 5 :	484	CONTROLES D'IDENTITE INDISPENSABLES	0.406
Moyenne :	3.404	PAS SE PLAINDRE SI TRAVAIL	0.353
Ec.type :	1.435	AIDER LES PAYS SOUS-DEVELOPPES	-0.353
c.c.m. :	0.741	POUR L'AUTO-DEFENSE	0.353
		ISOLER MALADES DU SIDA	0.351
		TROP DE GREVES	0.332

**\*\* 16 : LIMITER LES HERITAGES**

HISTOGRAMME		PRINCIPALES CORRELATIONS	
Modalité 1 :	625	EGALISER LES REVENUS	0.219
Modalité 2 :	447	LIMITER LE DEVELOPPEMENT DE L'INDUSTRIE	0.182
Modalité 3 :	233	TROP DE CHARGES SUR LES ENTREPRISES	-0.154
Modalité 4 :	100	ETUDIANTS PARASITES	0.142
Modalité 5 :	65	ORDINATEURS MENACENT NOS LIBERTES	0.130
Moyenne :	2.002	EXPORTER LES ARMES FRANCAISES	-0.129
Ec.type :	1.120	AIDER LES PAYS SOUS-DEVELOPPES	0.127
c.c.m. :	0.500	SE PASSER DE VOITURE	0.126
		CONTRE DISNEYLAND	0.124
		CENSURER CERTAINS LIVRES	0.123

**\*\* 18 : CONTRE LA PORNOGRAPHIE**

HISTOGRAMME		PRINCIPALES CORRELATIONS	
Modalité 1 :	253	CENSURER CERTAINS LIVRES	0.545
Modalité 2 :	268	POUR DES FEMMES NUES A LA TELE	-0.522
Modalité 3 :	292	LIBERALISER L'AVORTEMENT	-0.374
Modalité 4 :	251	DIEU EXISTE	0.359
Modalité 5 :	406	POUR LA FAMILLE	0.352
Moyenne :	3.197	HOMOSEXUELS COMME LES AUTRES	-0.347
Ec.type :	1.452	CONTROLES D'IDENTITE INDISPENSABLES	0.318
c.c.m. :	0.735	HASCHISCH EN VENTE LIBRE	-0.311
		RESPECTER LES CONVENANCES	0.310
		TROP DE GREVES	0.308

**\*\* 19 : PRIS POUR ABRUTIS A LA TELEVISION**

HISTOGRAMME		PRINCIPALES CORRELATIONS	
Modalité 1 :	103	ON N'APPREND PLUS RIEN A L'ECOLE	0.216
Modalité 2 :	279	GOVERNEMENT INEFFICACE	0.213
Modalité 3 :	412	PROGRES SCIENTIFIQUE NEFASTE	0.205
Modalité 4 :	325	HOMMES POLITIQUES INTEGRES	-0.182
Modalité 5 :	351	CONFIANCE EN LA JUSTICE	-0.180
Moyenne :	3.369	ORDINATEURS MENACENT NOS LIBERTES	0.177
Ec.type :	1.229	LA POLICE REMPLIT SA MISSION	-0.170
c.c.m. :	0.490	MEDECINE GRATUITE	0.165
		CONFIANCE AUX PSYCHIATRES	-0.155
		CONTRE DISNEYLAND	0.153

**\*\* 21 : AIDER LES PAYS SOUS-DEVELOPPES**

HISTOGRAMME		PRINCIPALES CORRELATIONS	
Modalité 1 :	173	TROP DE TRAVAILLEURS IMMIGRES	-0.353
Modalité 2 :	304	RETABLIR LA PEINE DE MORT	-0.330
Modalité 3 :	347	VOTE MUNICIPAL AUX ETRANGERS	0.320
Modalité 4 :	351	PAS DE PERMISSION DE SORTIE	-0.247
Modalité 5 :	295	MARCHE COMMUN FAVORABLE A LA FRANCE	0.203
Moyenne :	3.198	EXPORTER LES ARMES FRANCAISES	-0.203
Ec.type :	1.296	SYNDICATS INDISPENSABLES	0.198
c.c.m. :	0.581	ISOLER MALADES DU SIDA	-0.192
		HOMOSEXUELS COMME LES AUTRES	0.177
		L'EUROPE NE MARCHERA JAMAIS	-0.174

Enquête Agoramétrie 1987, n=1470 individus.

FIGURE 4.7 : Extrait de présentation rapide  
d'une matrice de corrélation (96x96)

La lecture simultanée de tous ces éléments permet :

- d'éliminer de l'ACP les variables ayant un histogramme très dissymétrique ; dans ce cas, on sait que les corrélations mesurées avec la variable sont sujettes à caution,

- de repérer des blocs de variables fortement corrélées entre elles, et ainsi d'éliminer certaines variables jugées trop redondantes,

- d'éliminer les variables qui, soit ont une petite liaison avec l'ensemble des autres variables (valeur faible du c.c.m.), soit sont trop redondantes par rapport au reste (valeurs élevées du c.c.m.). Dans les enquêtes traitées le coefficient de corrélation multiple  $R(x^j; (x^i/i \neq j))$ , pour les variables retenues, varie entre 0.4 et 0.7.

Rappelons que tous les coefficients de corrélation multiple peuvent être obtenus simultanément en inversant la matrice des corrélations : si  $r_{jj}$  est le  $j$ -ème terme diagonal de la matrice inverse, alors  $R(x^j; (x^i/i \neq j)) = [1/(1-r_{jj})]^{1/2}$ .

#### 4.4.2. Choix du nombre d'axes à retenir

Pour décider du nombre d'axes à retenir, on a le plus souvent recours à des critères empiriques, au moins en partie, : règle du coude, règle de Catell (point d'inflexion), règle de Guttman-Kaiser, règle du bâton cassé, test de sphéricité de Velicer, techniques de validation croisé... (voir, par exemple, [Joliffe 1986, ch. 6.1]).

La règle adoptée dans nos enquêtes d'opinion tranche avec ces méthodes empiriques qui fonctionnent, on ne sait pas très bien pourquoi ; elle repose sur l'interprétabilité des axes factoriels. Disposant d'un nombre élevé de variables qualitatives illustratives (signalétique) devant participer à l'interprétation des axes factoriels, nous retenons seulement les composantes factorielles présentant plusieurs liaisons significatives (au sens du  $F$ ) avec ces variables illustratives<sup>1</sup>

---

<sup>1</sup> Dans la pratique, pour les premières composantes principales, le nombre des liaisons significatives décroît, en gros, avec l'ordre de la composante. Le nombre d'axes à retenir est indiqué par une cassure (analogue à celle pour les valeurs propres avec la règle du coude).

Des analyses de la variance à un facteur sont alors effectuées de façon systématique. Pour donner un ordre de grandeur, dans le cas des enquêtes Agoramétrie, on procède à près de trois à quatre mille analyses de la variance : on analyse sur les vingt premières composantes l'influence de quelque deux cents variables du signalétique. Avec la procédure "à un seul passage" retenue, le temps d'exécution est à peine de quelques minutes. La valeur de la probabilité de dépassement de la statistique  $F$  est prise comme indicateur de l'intensité de la liaison. Il importe, comme pour le khi-2, que cette probabilité soit calculée avec une grande précision pour obtenir un rangement cohérent des liaisons par catégories d'intensité. Le recours au  $F$  de Fisher, malgré une certaine dépendance entre observations, est conforté ici, par la forme remarquablement régulière des histogrammes des premières composantes principales (voir figure 4.9).

Dans l'exemple de la figure 4.10 on est conduit, avec notre règle, à retenir sept composantes principales ; ce nombre est voisin de ce que l'on obtient avec les règles du coude ou de Catell appliquées sur l'histogramme des valeurs propres de la figure 4.8.

L'adoption de cette règle d'interprétabilité est en harmonie avec la façon de procéder en classification automatique (voir section 4.6). La classification s'effectue directement sur les composantes principales significatives, les classes de la typologie finale devant être interprétables à l'aide des variables du signalétique.

#### Remarque sur le calcul du $F$

Le calcul de la probabilité de dépassement  $\Pr\{F|\nu_1, \nu_2\}$  de la statistique  $F$  de Fisher est très précis quand le premier des nombres de degrés de liberté associés  $\nu_1$  est petit et pair (voir formule 26.6.4 dans [Zelen et al. 1964]). Ici, dans les analyses de la variance à un facteur,  $\nu_1$  est toujours petit : il est égal au nombre des modalités de la variable-facteur moins un. Donc, si  $\nu_1$  est pair, la formule exacte est appliquée ; sinon, on procède à une interpolation logarithmique analogue à celle décrite précédemment pour le khi-2 (section 4.3). Un calcul plus précis dans tous les cas de figure, mais aussi plus coûteux, est proposé dans [Laumon et al. 1985].

	Valeur propre	%	%Cumulé
1	9.429	9.82	9.82
2	5.949	6.20	16.02
3	4.129	4.30	20.32
4	3.133	3.26	23.58
5	2.729	2.84	26.43
6	2.371	2.47	28.90
7	1.779	1.85	30.75
8	1.561	1.63	32.38
9	1.470	1.53	33.91
10	1.435	1.49	35.40
11	1.317	1.37	36.77
12	1.286	1.34	38.11
13	1.275	1.33	39.44
14	1.231	1.28	40.72
15	1.222	1.27	42.00
16	1.187	1.24	43.23
17	1.181	1.23	44.46
18	1.140	1.19	45.65
19	1.125	1.17	46.82
20	1.094	1.14	47.96

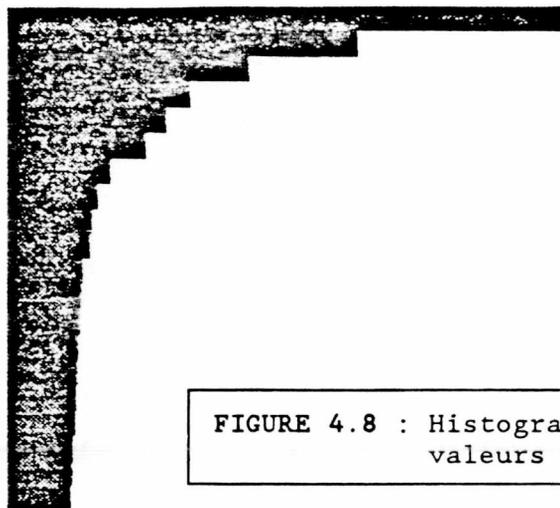


FIGURE 4.8 : Histogramme des valeurs propres

ENQUETE AGORAMETRIE 1987  
1470 Individus, 96 variables

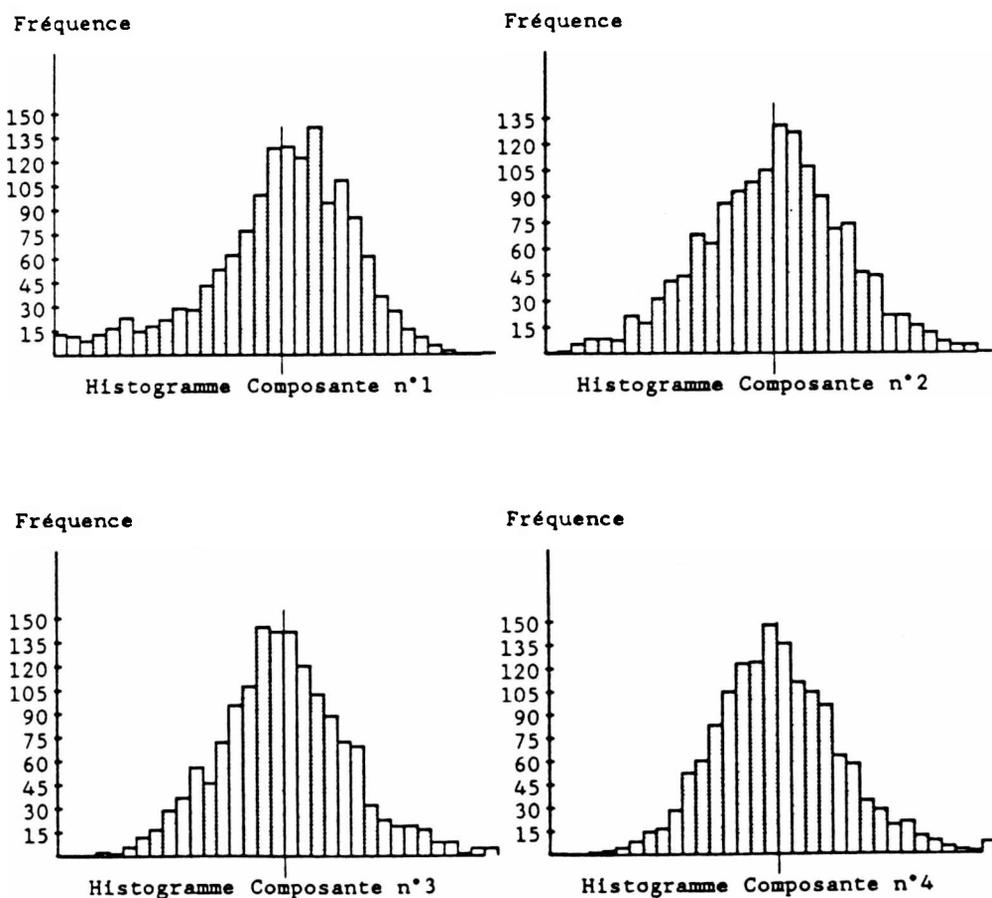


FIGURE 4.9 : Histogramme des 4 premières composantes principales

VARIABLE	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	TOTAL
REGION	■	■	■			■	■			■	■	■							7
HABITAT (4 MODALITES)	■	■				■				■	■							■	5
TYPE DE LOGEMENT	■	■	■			■	■			■	■	■							5
SEXE		■	■	■	■		■	■											5
AGE (5 MODALITES)	■		■		■	■	■					■	■						6
ETAT CIVIL	■		■		■		■					■							5
NOMBRE D'ENFANTS A CHARGE																			0
RELIGION	■		■	■		■		■	■		■								7
DIPLOME (7 MODALITES)	■	■				■	■												3
REVENU		■			■						■								3
CSP INTERVIEWE (8 MODALITES)	■	■	■	■	■		■	■		■									8
CSP INTERVIEWE (11 MODALITES)	■	■	■	■	■		■	■		■									8
CHOMAGE INTERVIEWE	■																		1
CSP CHEF DE FAMILLE (8 MODALITES)	■	■	■	■	■	■	■	■		■				■					9
CSP CHEF DE FAMILLE (11 MODALITE)	■	■	■	■	■	■	■	■		■				■					9
INTERVIEWE CHEF DE FAMILLE				■			■	■											3
POLITIQUE	■	■	■				■	■	■										5
PROXIMITE POLITIQUE	■	■	■		■	■	■	■					■						7
PERSONNALITE POLITIQUE PREFEREE	■	■	■	■	■	■	■	■											7
C.F.D.T.	■	■	■				■	■	■										5
C.G.C.		■	■	■				■	■										4
C.G.T.	■	■	■	■	■	■		■											6
F.O.	■	■	■	■				■									■		5
C.N.P.F.	■	■						■											3
R.BARRE	■	■			■	■	■												4
J.CHABAN DELMAS	■	■	■	■															3
J.P.CHEVENEMENT	■	■	■	■			■	■											4
J.CHIRAC	■	■																	2
J.DELORS		■	■	■			■	■											3
L.FABIUS	■	■	■	■		■	■												5
V.G.D'ESTAING	■	■																	1
L.JOSPIN	■	■	■	■				■											4
B.LALONDE	■	■	■							■				■					4
F.LEOTARD	■	■				■		■											4
J.M. LE PEN	■	■	■	■	■				■	■									5
G.MARCHAIS		■	■	■	■														2
F.MITERRAND	■	■	■	■	■														4
M.ROCARD	■	■	■	■			■	■											4
S.VEIL	■	■	■	■		■													4
PAYS LE PLUS FAVORABLE A LA FRAN		■	■				■	■											2
F.GONZALES		■	■																2
M.GORBATCHEV		■	■	■		■	■	■		■					■				7
H.KHOL	■	■	■	■				■											4
F.MITERRAND	■	■	■	■	■			■											4
R.REAGAN	■	■	■		■	■	■									■			6
M.THATCHER	■	■			■	■				■									4
CONTRE CANDIDAT SI DESACCORD NUC		■	■																1
MAUX DE TETE		■	■				■												2
MAL AU DOS		■	■																1
NERVOSITE		■	■																1
ETAT DEPRESSIF		■	■																1
INSOMNIE		■	■										■						3
VOITURE		■	■	■		■		■											5
LOGEMENT	■	■	■				■					■							4
RESIDENCE SECONDAIRE		■	■																2
PIANO		■	■	■															2
MAGNETOSCOPE		■	■	■				■						■					4
MICRO-ORDINATEUR		■	■			■													2
LECTEUR DISQUE COMPACT		■	■																1
PLAN EPARGNE LOGEMENT		■																	1
ACTIONS, OBLIGATIONS, SICAV	■	■	■		■														3
BIENS IMMOBILIERS DE RAPPORT	■	■	■	■															3
USAGE DU MINTEL	■	■	■																2
ACHAT ACTIONS ENT.PRIVATISEE	■	■	■		■														3
SECURITE SOCIALE	■	■	■	■	■			■											5

FIGURE 4.10 : Bilan des liaisons entre variables du signa-  
létique et composantes principales (Suite...)

VARIABLE	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	TOTAL
AUCUN QUOTIDIEN		■																	1
L'EQUIPE																			0
FRANCE SOIR																			0
LE FIGARO	■■■	■■■																	2
L'HUMANITE	■																		1
LIBERATION	■	■																	1
LE MATIN	■																		1
LE MONDE	■	■	■■■																2
QUEST FRANCE											■								1
LE PARISIEN LIBERE				■															1
LE QUOTIDIEN DE PARIS																			0
LA CROIX				■		■													2
QUOTIDIENS ECONOMIQUES		■	■																1
AUTRES NATIONAUX																			0
AUTRES REGIONAUX	■	■				■													2
AUCUN MAGAZINE		■	■																1
ACTUEL	■■■																		1
ALMA																			0
PARENTS																			0
CANARD ENCHAINE	■	■																	1
COSMOPOLITAN																			0
ELLE							■	■											1
L'EVENEMENT DU JEUDI	■	■																	1
L'EXPANSION		■	■																1
L'EXPRESS		■■■																	1
FEMME ACTUELLE						■	■												1
FEMME D'AUJOURD'HUI																			0
FIGARO MAGAZINE	■■■	■	■																2
JOURS DE FRANCE																			0
MARIE-CLAIRE																			0
MARIE-FRANCE																			0
BANDES DESSINEES	■	■																	1
MODES ET TRAVAUX																			0
LE NOUVEL OBSERVATEUR	■																		1
PARIS MATCH	■																		1
MINUTE																			0
LE POINT																			0
PRIMA																			0
LIRE																			0
SCIENCE ET VIE															■				1
SCIENCES ET AVENIR																			0
V.S.D.			■																1
LUI																			0
ICI-PARIS,...		■																	1
FOYER, MODE, BRICOLAGE																			0
MAGAZINES ECONOMIQUES																			0
MAGAZINES D'INFORMATIQUE																			0
AUTOMOBILES		■																	1
MAGAZINES CATHOLIQUES			■	■															1
MAGAZINES DE CONSOMMATION																			0
MAGAZINES DE SANTE																			0
PRESSE DU COEUR																			0
AUTRES MAGAZINES																			0
TOTAL	96	99	51	29	41	27	41	11	7	9	1	8	11	1	4	0	1	4	441

Significativité des liaisons :

■ : à  $10^{-3}$ ■■ : à  $10^{-4}$ ■■■ : à  $10^{-6}$ ■■■ : à  $10^{-12}$ ■■■■ : à  $10^{-30}$ 

LISTE DES ABREVIATIONS : C1...C18 : 18 premières composantes principales

Enquête Agoramétrie 1987, n=1470 individus.

FIGURE 4.10 : (...Suite)

#### 4.5. Analyse des Correspondances

Si on aborde l'opinion à l'aide d'un questionnaire très homogène (propositions référendum avec réponses sous forme d'une même échelle d'accord), l'ACP peut servir, de façon quasi-exclusive, comme méthode d'analyse multidimensionnelle. Le logiciel comprend toutefois un module d'Analyse Factorielle des Correspondances Multiples qui permet, au delà de l'analyse multiple usuelle, d'effectuer une analyse "optimale".

Notons

- $x^1, \dots, x^m$ , un ensemble de  $m$  variables qualitatives,
- $X_i (p_i \times n)$ , le tableau des indicatrices associé à  $x^i$ ,
- $X (q \times n)$  ;  $q = \sum_{i=1}^m p_i$ , le tableau disjonctif complet.  $X' = [X_1' | \dots | X_m']$ ,
- $D_p (n \times n)$ , la métrique diagonale des poids,
- $B (q \times q)$ , le tableau de Burt.  $B = X D_p X'$ ,
- $D_p (q \times q) = \text{diag}(B)$ , la matrice diagonale des probabilités marginales

Dans l'analyse usuelle (analyse des Correspondances du tableau  $X$  ; matrice à diagonaliser :  $D_{1/p} B$ ), nous avons recours aux normalisations et algorithmes proposés par [Lebart et al. 1984].

L'analyse "optimale" a été introduite par [Cazes et al. 1976]. Il s'agit d'une application de la méthode STATIS (cf 3.1.2.) au cas des variables qualitatives. Avant de traiter le tableau disjonctif complet, chaque tableau  $X_i$  est pondéré par un coefficient issu d'une analyse des dépendances entre variables qualitatives prises globalement.

Soit

- $t_{ij} = \frac{\phi^2(x^i, x^j)}{[(p_i - 1)(p_j - 1)]^{1/2}}$ , le coefficient de Tschuprow entre les variables  $x^i$  et  $x^j$  (cosinus de l'angle entre les opérateurs de projection associés),
- $T (m \times m)$ , la matrice de terme général  $t_{ij}$  (équivalente à une matrice de corrélation dont tous les termes seraient positifs),
- $\alpha (m \times 1)$ , le premier vecteur propre de  $T$ , dont toutes les coordonnées  $\alpha_1, \dots, \alpha_m$  sont prises positives,

$$- a_i = \frac{\alpha_i}{(p_i - 1)^{1/2}} ; i=1, \dots, m$$

$$- \tilde{X}' = [a_1 X_1' | \dots | a_m X_m']$$

Chaque tableau  $X_i$  étant pondéré par le coefficient  $a_i$ , l'analyse optimale revient à une analyse des correspondances effectuée, non pas sur  $X$ , mais sur le tableau  $\tilde{X}$  ; les variables bien liées ( $a_i$  élevé) se trouvent alors largement favorisées. La matrice à diagonaliser n'est autre que :  $D_{1/p} B D_a$ , où  $D_a (q \times q)$  est une matrice diagonale où chacun des  $a_i$  apparaît  $p_i$  fois. Pour une discussion sur les propriétés d'optimalité de cette analyse des correspondances particulière, on consultera [Saporta 1979].

Parmi les informations fournies par le programme, on trouve : les contributions absolues et relatives, les poids  $a_i$  (cas d'une analyse optimale) et, si on le demande, la matrice  $T$  (dans un fichier) des coefficients de Tschuprow. Avec cette dernière matrice, équivalente à une matrice de corrélation, on pourra opérer une classification des variables  $x^i$  prises globalement (voir section 4.8, plus loin).

#### 4.6. Classification automatique

Ce module permet d'effectuer une classification des individus à partir d'un fichier de composantes factorielles issues d'une ACP ou d'une AFC ou, plus généralement, d'un fichier de variables continues comme, par exemple, les composantes d'une "greffe". La procédure de classification est celle proposée dans le logiciel SPAD [Lebart et al 1985] ; nous l'avons reprise en améliorant son efficacité au niveau des calculs et de la présentation des résultats.

Rappelons en les principales étapes :

- Dans une première étape, on utilise une version simplifiée de la méthode des nuées dynamiques. Partant d'un tirage au hasard de  $k$  individus (la valeur par défaut  $k=8$ , est adéquate pour des enquêtes d'envi-

ron 1000 individus), considérés au départ de l'algorithme comme centres des classes, on affecte les individus au centre le plus proche. Les centres de gravité des classes définissent alors de nouveaux centres autour desquels les individus sont réaffectés... Le nombre d'itérations (par défaut 5) est fixé à l'avance. L'opération est répétée  $m$  (par défaut 2) fois ; puis on établit le croisement entre les  $m$  partitions ainsi obtenues. Parmi les  $k^m$  classes produites, on retient les  $l$  (par défaut 30) classes ayant les plus forts effectifs ; les classes restantes sont regroupées dans une même classe "fourre-tout".

- dans une deuxième étape, pour regrouper les  $l+1$  classes précédemment définies, on procède à une classification ascendante hiérarchique (algorithme du plus proche voisin utilisant, par défaut, la distance de Ward).

- dans la troisième étape, on produit des "aides à l'interprétation" de la partition finale considérée comme une nouvelle variable qualitative : tableaux croisés avec toutes les autres variables jugées intéressantes ; projection des centres de gravité des modalités-classes sur les axes factoriels. Pour aller plus vite dans l'examen des classes de la partition, on fournit un classement par ordre d'importance, relativement à la caractérisation de la classe, des variables continues, d'une part, et des modalités de la totalité des variables qualitatives, d'autre part. Pour cela on s'appuie sur la procédure des "valeurs-tests" [Morineau 1984] : on ordonne les variables continues (les modalités des variables qualitatives) en fonction des résultats du test d'hypothèse d'égalité des moyennes (des pourcentages) effectué dans la classe puis dans l'ensemble des individus.

Travailler sur les composantes principales pour la classification signifie "élimination du bruit". Combien de composantes faut-il alors considérer ? Une réponse est de retenir seulement les composantes qui sont significativement liées aux variables du signalétique dans la mesure où les classes de la typologie finale sont interprétées à l'aide des modalités de ces mêmes variables. Cette démarche est cohérente avec la procédure retenue pour le choix du nombre d'axes (paragraphe 4.4.3).

#### 4.7. Rotations Analytiques

Avec ce module, on met en évidence, s'ils existent, les groupes homogènes de variables.

Les rotations analytiques sont un ensemble de techniques qui ont été développées autour du modèle factoriel classique "en facteurs communs et spécifiques" pour permettre l'obtention d'indicateurs factoriels plus interprétables. Partant d'une solution factorielle initiale (ici on prendra celle de l'ACP), on cherche un ensemble de combinaisons linéaires des composantes factorielles initiales, tel que chaque variable puisse être expliquée par un nombre moindre de composantes.

Rappel théorique :

Soit  $C(k \times n)$  le tableau des  $k$  premières composantes principales normées extraites de l'ACP du triplet  $(X, M, D_p)$  ; elles forment une base orthonormée dans  $F_1 = R^k$  (pour simplifier, le tableau des données  $X(p \times n)$  est pris ici centré-réduit). Soit encore  $U(p \times k)$  la matrice des corrélations entre variables et composantes factorielles :  $U = XD_p C'$  ;  $U$  est aussi la matrice des coefficients de régression ("saturations") de  $X$  par rapport à  $C$  :

$$X = UC + E \quad (4.3)$$

Dans cette écriture chaque variable de  $X$  est considérée comme étant "expliquée" par les  $k$  composantes factorielles. Remarquons que (4.3) peut se réécrire :  $X = UTT^{-1}C + E = U_R C_R + E$ , où  $T(k \times k)$  est une matrice de "rotation" associée au changement de base :  $C \rightarrow C_R$ , dans  $R^k$  ;  $C_R = T^{-1}C$ . Quelle que soit la base  $C_R$  choisie dans  $R^k$ , l'approximation  $\hat{X} = UC = U_R C_R$  de  $X$  fournie par les variables de la base reste inchangée. Pour l'interprétabilité des composantes factorielles on a alors intérêt à choisir une base "optimale"  $C_R$  de sorte que :

(i) chaque variable  $x^i$  soit expliquée par un nombre moindre de composantes factorielles (autant de zéros que possible dans les lignes de la matrice  $U_R$ ) ; on parle alors de solution "parcimonieuse",

(ii) chaque composante soit explicative d'un nombre moindre de variables (autant de zéros que possible dans les colonnes de la matrice  $U_R$ ) ; ici, on parle de solution "simple".

Avec les techniques de rotation analytique, on cherche la matrice de rotation  $T$ , sous la contrainte de préservation de la longueur unitaire des composantes factorielles (i.e.  $\text{diag}(T'T)^{-1} = I$ ), de sorte que la matrice  $V=UT$  soit aussi "simple" que possible. Pour cela on utilise différents critères qui reviennent tous à la minimisation du critère de simplicité de Harman :

$$B = \sum_{\ell=1}^k \sum_{m=1}^k \left[ \sum_{j=1}^p v_{j\ell}^2 v_{jm}^2 - \frac{\gamma}{p} \left( \sum_{j=1}^p v_{j\ell}^2 \right) \left( \sum_{j=1}^p v_{jm}^2 \right) \right] ; j \neq m \quad (4.4)$$

De petites valeurs de gamma ( $\gamma=0$ ) privilégient la simplicité des lignes de  $V$  et des valeurs élevées ( $\gamma=1$ ) privilégient la simplicité des colonnes de  $V$ .

La qualité des résultats est souvent améliorée si, avant la rotation, la normalisation de Kaiser est appliquée à  $U$  : Si, pour  $i=1, \dots, p$  on note  $h_i^2 = \sum_{j=1}^k u_{ij}^2$ ,  $D_h(p \times p)$  étant la matrice diagonale constituée des  $h_i$ , la minimisation de (4.4) s'effectue maintenant ayant remplacé  $U$  par la matrice normalisée  $U^* = D_h^{-1} U$ .

Dans le cas particulier où  $T$  est orthogonale :  $T'T = I$ , la rotation est aussi appelée orthogonale ; si, de plus,  $\gamma=1$ , on trouve la rotation "Varimax" qui est la plus souvent utilisée dans la pratique. Ici, les différents types de rotation orthogonale<sup>1</sup> ont été programmés sur la base de l'algorithme itératif de Horst et Sherin [Sherin 1966] ; à chaque itération, la matrice  $T$  est calculée directement à partir de la matrice initiale  $U$  (algorithme autocorrectif). Il est connu (voir [Ten Berge 1984]) que cet algorithme peut conduire à des optima locaux et qu'il est moins rapide que l'algorithme initial de Kaiser (rotation itérative par couple de composantes) qui, cependant n'est pas autocorrectif. Si donc nous avons opté pour le premier algorithme, c'est pour éviter les erreurs d'arrondi qui peuvent survenir lorsqu'on traite des matrices  $U$  de grande dimension ; dans ce cas, un nombre élevé d'itérations est souvent nécessaire. Avec notre procédures, par ailleurs, la "rotation" des matrices  $U$  de grande dimension (100x15, par exemple) nécessite quelques minutes seulement.

---

<sup>1</sup>qui correspondent à différentes valeurs de  $\gamma$  et qui sont connus dans la littérature sous les noms : "quartimax" pour  $\gamma=0$ , "equamax" pour  $\gamma=p/2$ , "parsimax" pour  $\gamma=p(k-1)/(p+k-2)$ ...

Quand  $T$  n'est pas orthogonal, on parle (abusivement) de "rotations" obliques. Pour traiter de ces cas, nous avons implanté deux algorithmes. Le premier, "Promax" [Hendrickson et White 1964] conduit très rapidement, à partir d'une solution orthogonale, à une solution oblique très proche de l'optimum ; grâce à son efficacité, il est souvent préféré du deuxième "Direct quartimin" [Jennrich et Sampson 1966] qui fournit une solution de meilleure qualité. Ce dernier algorithme est aussi le choix recommandé dans BMDP.

A la place de la matrice des coefficients de régression  $U_R$  (appelée "factor pattern"), on peut éventuellement rechercher la simplicité de la matrice des corrélations entre variables et composantes ("factor structure")  $W=U_R(C_R D_p C_R')$ . Quand la solution est orthogonale les deux matrices coïncident ; quand elle est oblique, on doit toujours minimiser (4.4) (avec  $V=WT$ ) mais cette fois sous la contrainte  $\text{diag}(T'T)=I$ . Pour traiter de ce cas, nous nous sommes appuyé sur l'algorithme de Harman [Harman 1967].

Tous les algorithmes étant itératifs, les procédures informatiques s'arrêtent, soit quand le nombre d'itérations excède une limite introduite par l'utilisateur, soit quand la différence entre deux valeurs successives de  $B$  dans (4.4) devient inférieure à une valeur introduite aussi par l'utilisateur. La valeur de  $B$  étant par ailleurs affichée à l'écran, on peut encore arrêter les itérations durant l'exécution en jugeant la convergence lente.

Soulignons ici que, par la construction de nos questionnaires, nous n'avons pratiquement jamais besoin à faire appel à des techniques de rotation pour pouvoir interpréter les indicateurs factoriels. Les rotations effectuées ont alors un but classificatoire : un groupe de variables bien expliquées par une même composante doit être relativement homogène.

Pour mettre en évidence les groupes, les composantes de la matrice finale  $V$  sont présentées par ordre décroissant de leur pouvoir explicatif. En même temps, à l'intérieur de chaque composante, les variables ayant les plus fortes saturations sont triées (voir figure 4.11).

Radiographie	<b>-0.61</b>	-0.21	-0.05	-0.01	0.11	-0.04	-0.03	0.15
Transfusions Sanguines	<b>-0.57</b>	-0.11	-0.25	-0.05	0.13	0.02	-0.07	0.04
Anesthésies	<b>-0.54</b>	-0.26	-0.19	-0.12	0.10	0.01	-0.10	-0.00
Antibiotiques	<b>-0.52</b>	-0.01	-0.12	-0.09	0.19	-0.05	-0.00	0.25
Aspirine	<b>-0.52</b>	-0.26	-0.05	-0.09	0.18	0.01	-0.02	0.26
Vaccins	<b>-0.47</b>	0.14	-0.19	0.02	-0.01	-0.13	0.03	0.32
Foot-Ball	<b>-0.46</b>	-0.03	-0.01	-0.02	0.07	-0.33	0.09	0.10
Engrais	<b>-0.37</b>	-0.15	-0.00	-0.29	0.18	-0.05	-0.29	0.22
Eau de Javel	-0.05	<b>-0.64</b>	-0.03	-0.12	0.02	0.02	-0.14	0.12
Prises de Courant	0.09	<b>-0.61</b>	-0.26	-0.05	0.29	-0.04	-0.17	0.17
Cocotte Minute	-0.34	<b>-0.56</b>	-0.18	-0.03	0.09	-0.03	0.02	0.22
Tondeuses à Gazon	-0.25	<b>-0.56</b>	-0.12	0.01	0.07	-0.19	-0.14	0.03
Alpinisme	-0.19	<b>-0.51</b>	-0.17	-0.19	-0.11	-0.22	-0.17	-0.14
Téléviseur	-0.16	<b>-0.45</b>	-0.29	-0.00	0.30	-0.04	0.08	0.20
Chasse	-0.23	<b>-0.41</b>	-0.05	-0.28	0.18	-0.10	-0.09	0.06
Armes à Feu	0.20	<b>-0.40</b>	0.05	-0.29	<b>0.36</b>	-0.07	-0.19	-0.04
Jouets Electriques	-0.07	<b>-0.39</b>	-0.28	-0.04	0.32	-0.13	-0.06	0.20
Mines de Charbon	-0.27	-0.30	-0.26	-0.28	0.05	-0.19	-0.06	-0.09
Avion	-0.06	-0.10	<b>-0.70</b>	0.01	0.06	-0.09	-0.05	0.08
Train	-0.11	-0.05	<b>-0.69</b>	0.04	0.04	-0.13	0.04	0.02
Barrages Hydroélectriques	-0.15	-0.24	<b>-0.55</b>	-0.24	0.13	-0.10	0.04	0.12
Rayon Laser	-0.19	-0.22	<b>-0.53</b>	-0.23	0.08	-0.01	-0.06	0.10
Médecine Nucléaire	-0.03	-0.11	<b>-0.49</b>	-0.24	0.28	0.06	0.01	0.15
Ski	-0.29	-0.27	-0.33	-0.07	-0.02	-0.31	-0.18	-0.09
Dechets Nucléaires	0.01	-0.15	0.05	<b>-0.72</b>	0.13	0.00	-0.10	0.12
Dechets Chimiques	-0.13	-0.10	-0.03	<b>-0.70</b>	0.15	0.03	-0.14	0.05
Centrales Nucléaires	-0.15	-0.05	-0.32	<b>-0.63</b>	0.01	0.03	-0.07	-0.02
Industrie Chimique	-0.13	0.09	-0.20	<b>-0.57</b>	0.14	-0.24	0.03	0.08
Bombe Atomique	0.12	-0.13	0.03	<b>-0.56</b>	0.02	-0.00	-0.19	0.06
Raffineries de Petrole	-0.24	0.02	-0.34	<b>-0.46</b>	0.05	-0.22	0.00	0.09
Somnifères	-0.13	-0.17	-0.12	-0.06	<b>0.64</b>	0.04	-0.32	0.03
Gaz d'Echappement	0.03	-0.20	-0.05	-0.17	<b>0.57</b>	-0.12	-0.09	0.25
Tranquillisants	-0.25	-0.12	-0.02	-0.11	<b>0.56</b>	-0.01	-0.32	-0.00
Manipulations Génétiques	-0.14	0.02	-0.16	-0.09	<b>0.49</b>	-0.13	-0.07	-0.08
Travail sur Ecran	-0.23	-0.09	-0.25	0.00	<b>0.48</b>	0.00	0.07	0.15
Plomb dans l'Essence	-0.29	-0.01	0.12	-0.21	<b>0.44</b>	-0.10	0.03	0.17
Mobyette	0.08	-0.10	0.09	0.09	0.09	<b>-0.70</b>	-0.17	0.19
Moto	0.05	-0.15	-0.05	-0.09	0.07	<b>-0.64</b>	-0.24	0.14
Poids Lourds	-0.11	0.04	-0.14	-0.14	-0.03	<b>-0.59</b>	-0.03	-0.01
Voiture	-0.08	-0.04	-0.12	-0.01	0.20	<b>-0.56</b>	-0.24	-0.16
Bicyclette	-0.13	<b>-0.38</b>	-0.03	0.04	0.01	<b>-0.47</b>	0.05	0.09
Industrie du Bâtiment	-0.27	-0.15	-0.25	-0.19	0.07	<b>-0.38</b>	0.10	0.04
Alcool	-0.01	-0.13	-0.00	-0.15	0.11	-0.18	<b>-0.71</b>	0.09
Drogue	0.08	-0.12	-0.01	-0.11	0.11	0.01	<b>-0.67</b>	-0.03
Tabac	-0.13	-0.08	0.03	-0.15	0.17	-0.13	<b>-0.66</b>	0.03
Vin	-0.03	-0.04	-0.07	-0.01	-0.00	-0.31	<b>-0.48</b>	0.14
Oranges Traitées	-0.12	-0.05	-0.02	-0.11	0.21	-0.12	-0.01	<b>0.67</b>
Conserves	-0.25	-0.17	-0.21	-0.13	-0.07	0.03	-0.05	<b>0.55</b>
Poulet Industriel	-0.29	-0.19	0.06	-0.18	0.22	-0.08	-0.04	<b>0.49</b>
Pilule	-0.30	-0.06	-0.27	0.01	0.01	-0.06	-0.22	<b>0.46</b>
Fours à Micro-ondes	-0.14	-0.28	-0.31	-0.01	0.09	-0.11	-0.04	<b>0.38</b>
Avortement	-0.20	-0.09	-0.34	-0.05	0.00	0.00	-0.28	<b>0.36</b>
<b>SOMME DES CORRELATIONS CARREES :</b>								
	3.48	3.43	3.25	3.11	2.71	2.68	2.52	2.34

Enquête sur la perception des risques (voir aussi, figure 4.12)

Rotation varimax des 8 premières comosantes principales de l'ACP sur les activités dangereuses. Apparaissent en gras sur fond grisé, les corrélations supérieures à 0.35 en valeur absolue.

FIGURE 4.11 : Rotation Varimax

#### 4.8. Classification de variables

Dans la majorité des programmes de classification automatique, l'utilisateur a la possibilité d'introduire une matrice des corrélations (les valeurs absolues des corrélations) en tant que matrice de similarités entre éléments-variables, et opérer une classification (ascendante hiérarchique, en principe) à partir de cette matrice. Cependant les techniques utilisées ("single linkage", "average linkage"...), s'appuient sur des indices de distance (ou de similarité) qui ne tiennent pas compte du fait que les unités à classer sont des variables.

Pour compléter les indices précédents, nous avons songé au coefficient RV (voir aussi section 3.2.3.), indice plus approprié pour mesurer la similarité entre deux ensembles de variables. Rappelons que, quand les variables centrées sont représentées par des vecteurs de  $R^n$ , le coefficient RV est une mesure de la proximité entre les sous-espaces engendrés par deux paquets de variables ; étant égal au cosinus entre les opérateurs de projection orthogonale sur les deux sous-espaces, il varie entre zéro (espaces disjoints) et un (espaces confondus).

Dans ce module, on procède à la classification ascendante hiérarchique en munissant l'ensemble de variables d'un indice de similarité basé toujours sur la matrice des corrélations. Au début, chaque variable forme une classe ; à chaque itération, les deux classes le plus proches sont fusionnées et on continue jusqu'à ce qu'il ne reste plus qu'une seule classe.

Trois indices de similarité ont été retenus :

$$- D_1(A,B) = \max_{i \in A, j \in B} |\text{cor}(x^i, x^j)| ; \text{ (single linkage)}$$

$$- D_2(A,B) = \frac{\sum_{i \in A} \sum_{j \in B} |\text{cor}(x^i, x^j)|}{pq} ; \text{ (average linkage)}$$

$$- D_3(A,B) = \frac{\text{trace} \left( \begin{matrix} R_{AA}^{-1} & R_{AB}^{-1} \\ R_{BA}^{-1} & R_{BB}^{-1} \end{matrix} \right)}{(pq)^{\frac{1}{2}}} = \frac{\sum_{i=1}^{\min(p,q)} \rho_i^2(A,B)}{(pq)^{\frac{1}{2}}} \text{ canonical linkage}$$

où p et q sont respectivement les effectifs des classes A et B, et

où  $\rho_i(A,B)$  est le  $i$ -ème coefficient de corrélation canonique entre les variables des classes A et B.

La présentation de l'arbre de classification est celle du programme PLM (Cluster analysis of variables) de BMDP : ayant multiplié par cent et arrondi par la suite les valeurs absolues des corrélations, on imprime la partie triangulaire supérieure (sans la diagonale) de la matrice réordonnée par la procédure de classification ; les classes  $y$  sont illustrées par des lignes horizontales ou obliques.

Dans l'arbre de la figure (4.12), on peut constater que la classification, avec l'indice  $D_3$ , des 52 variables concernant la perception des activités dangereuses est très cohérente. On pourra comparer cette typologie avec les résultats de la rotation varimax effectuée sur les huit premières composantes principales issues de la diagonalisation de la (52x52) matrice des corrélations correspondante (figure 4.11).

#### 4.9. Stabilité des points moyens

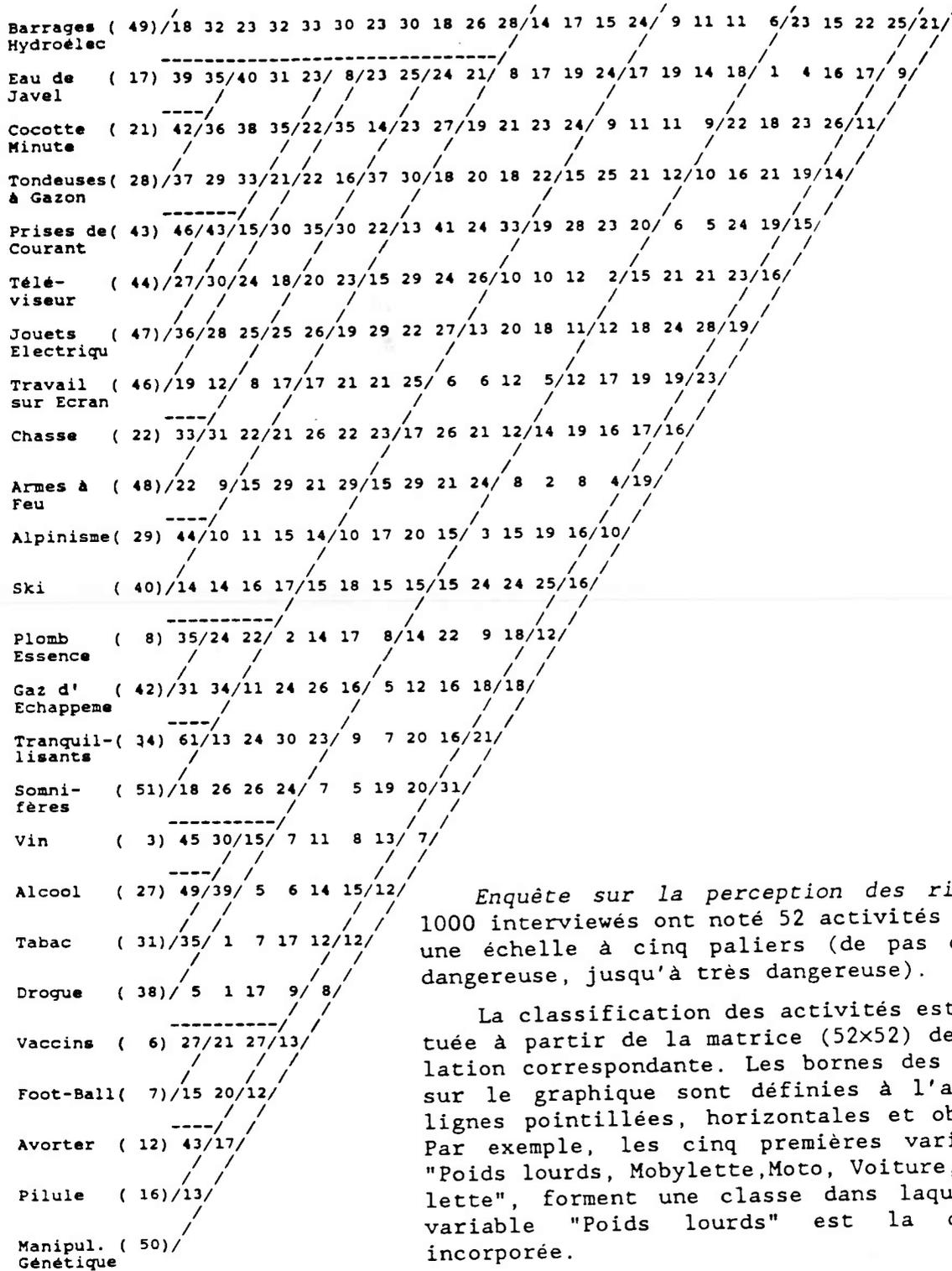
Dans les enquêtes d'opinion traitées, les variables qualitatives illustratives jouent un rôle important en analyse en composantes principales. Toutes les modalités de ces variables apparaissant au même titre sur un graphique (aussi bien celles à faible effectif que celles à effectif élevé), il a été jugé nécessaire d'intégrer dans le logiciel un module permettant un examen rapide de la stabilité des points moyens sur les plans principaux.

Pour cela on a été inspiré d'une proposition faite dans [Hatabian et Saporta 1985]. En étudiant l'utilisation des zones de confiance autour des points d'un graphique d'analyse factorielle, les auteurs de l'article font la distinction entre deux optiques :

- (i) inférence statistique. Considérant que les résultats obtenus sont des estimations des caractéristiques d'une "population", les auteurs s'appuient alors sur des hypothèses de normalité pour construire des ellipses de confiance autour des centres de gravité initiaux.

VARIABLE	N°																												
Poids Lourds	( 1)	27	28	25	15	33	6	6	5	5	8	9	10	6	8	12	12	12	9	16	20	16	14	19	5	15	13	15	2
Mobylette	( 4)	52	31	35	6	5	1	6	5	19	12	8	3	5	5	9	9	4	9	17	10	14	3	3	0	6	5	8	9
Moto	( 15)	36	27	13	17	12	10	11	17	15	16	8	13	9	11	14	5	20	25	18	14	9	9	11	13	19	18	13	23
Voiture	( 39)	25	13	7	9	12	7	4	9	12	6	10	5	9	11	9	15	20	19	11	10	8	13	17	8	6	15	17	13
Bicyclette	( 35)	6	5	7	8	2	10	17	14	11	13	15	17	16	14	11	16	24	30	15	10	11	10	18	17	25	25	17	22
Industrie Chimique	( 2)	34	37	34	17	16	15	25	14	20	18	9	16	15	36	22	21	15	25	21	15	15	22	10	15	13	12	16	16
Dechets Nucléaire	( 18)	52	38	41	17	18	24	14	10	17	12	17	8	25	8	17	11	16	20	2	4	17	21	15	10	18	15	15	11
Dechets Chimiques	( 26)	39	28	12	24	30	16	15	17	11	22	19	31	15	23	11	24	22	4	9	20	17	17	11	18	12	17	12	30
Centrales Nucléaire	( 30)	29	9	17	22	20	17	13	15	19	22	34	21	23	17	32	34	18	17	35	15	15	10	15	15	16	10	26	13
Bombe Atomique	( 10)	7	12	14	7	5	9	3	6	2	16	8	14	7	12	10	0	3	12	11	5	9	12	7	8	6	14	27	16
Oranges Traitées	( 5)	39	31	34	24	28	21	16	14	17	13	14	26	17	20	9	13	22	14	24	13	18	21	22	23	19	7	9	14
Poulet Industrie	( 24)	35	32	24	31	25	27	25	17	18	20	30	19	20	6	9	19	16	27	21	21	29	22	27	27	15	15	18	27
Engrais	( 32)	24	20	32	27	27	21	24	17	32	24	23	19	7	10	25	15	27	24	21	23	23	20	25	15	23	25	25	28
Conserves	( 11)	31	28	24	23	21	21	20	20	29	25	17	16	20	25	18	31	16	22	23	25	15	19	3	17	15	8	17	16
Antibiotiques	( 9)	37	31	26	33	25	26	18	19	25	16	17	18	17	15	26	19	14	22	18	23	19	4	14	14	28	20	26	24
Aspirine	( 19)	49	34	27	20	24	20	22	20	18	16	19	27	26	37	24	23	27	24	27	27	13	13	20	17	20	27	26	13
Radio-graphie	( 20)	36	30	21	21	17	24	23	19	12	18	23	17	37	26	16	27	16	26	22	5	16	23	21	13	20	20	12	9
Anesthésies	( 23)	41	20	21	29	23	28	24	20	23	29	17	32	23	22	26	24	22	34	9	25	21	17	12	25	23	6	10	18
Transfus. Sanguines	( 25)	25	20	24	27	30	22	22	16	23	13	27	27	18	30	21	23	20	2	17	21	20	16	20	18	12	11	14	2
Raffiner. Petrole	( 13)	39	32	17	28	23	22	27	35	15	21	20	14	18	17	15	18	9	15	23	22	19	17	14	10	15	10	6	16
Industrie Bâtiment	( 14)	31	17	26	12	24	18	29	17	23	26	22	19	28	19	17	7	19	23	14	18	12	12	13	9	9	1	16	25
Mines de Charbon	( 33)	18	32	22	20	23	34	20	26	23	26	25	22	17	24	15	32	28	14	18	24	18	10	17	19	9	8	15	17
Fours à Micro-ond	( 36)	38	27	23	23	27	18	35	29	24	31	28	26	19	7	18	20	14	18	17	18	13	12	17	7	14	12	22	26
Rayon Laser	( 37)	38	30	33	34	22	33	26	26	26	29	23	24	13	25	22	9	18	20	19	10	13	12	11	17	11	25	20	17
Médecine Nucléaire	( 45)	22	27	36	13	21	16	27	24	24	29	22	13	16	17	14	19	15	24	8	11	7	4	13	4	20	18	21	
Train	( 41)	47	34	8	23	14	19	28	21	17	12	2	14	34	7	15	5	7	8	4	4	2	17	14	18	18	7		
Avion	( 52)	41	9	24	18	24	27	21	17	17	9	21	27	5	13	12	17	11	7	5	1	16	8	23	23	6			
Barrages Hydroélec	( 49)	18	32	23	32	33	30	23	30	18	26	28	14	17	15	24	9	11	11	6	23	15	22	25	21				

FIGURE 4.12 : Classification sur Matrice de Corrélation (Suite...)



Enquête sur la perception des risques : 1000 interviewés ont noté 52 activités suivant une échelle à cinq paliers (de pas du tout dangereuse, jusqu'à très dangereuse).

La classification des activités est effectuée à partir de la matrice (52x52) de corrélation correspondante. Les bornes des classes sur le graphique sont définies à l'aide des lignes pointillées, horizontales et obliques. Par exemple, les cinq premières variables : "Poids lourds, Mobylette, Moto, Voiture, Bicyclette", forment une classe dans laquelle la variable "Poids lourds" est la dernière incorporée.

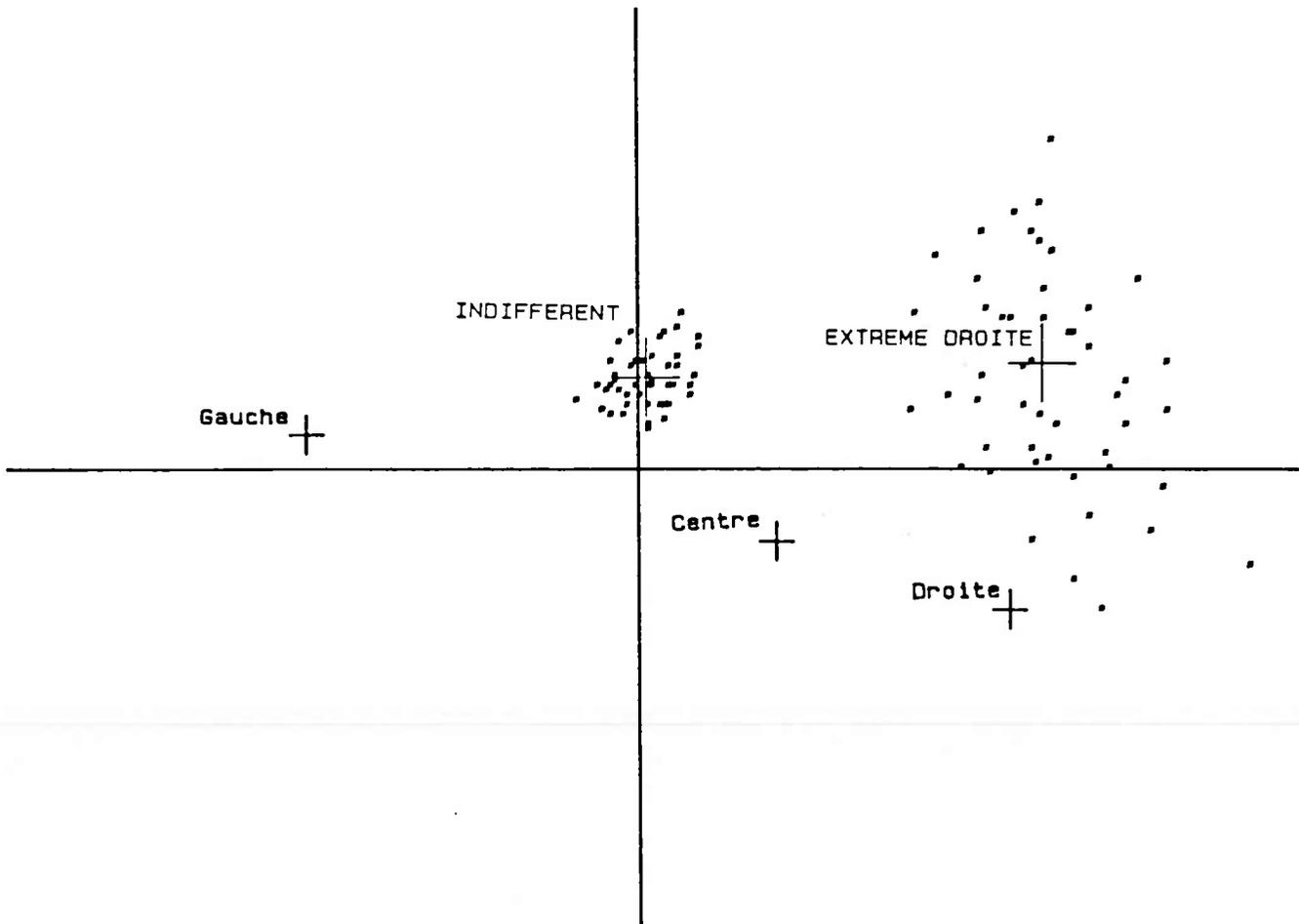
Les chiffres qui apparaissent sur l'arbre sont les valeurs absolues des corrélations multipliées par cent et arrondies (corrélation de 0.27 entre "Poids lourds" et "Mobylette", de 0.28 entre "Poids lourds" et "Moto", de 0.52 entre "Mobylette" et "Moto"...). Plus on va vers la droite de l'arbre, plus les corrélations sont faibles. C'est pourquoi la petite partie triangulaire manquante dans la page précédente, est non significative.

FIGURE 4.12 : (... Suite)

(ii) validité des résultats. Pour étudier la stabilité des points vis à vis de perturbations des données, les auteurs proposent, ici, d'avoir recours aux méthodes de rééchantillonnage (bootstrap, jackknife...)

C'est dans la deuxième optique, proche de nos préoccupations, que l'on s'est placé. A partir des coordonnées factorielles initiales, on procède à N tirages avec remise (répétitions bootstrap) à l'intérieur de chaque catégorie (modalité) étudiée d'une variable qualitative. La dispersion, sur les plans principaux, des N nouveaux points moyens ainsi produits permet d'apprécier la stabilité du point moyen de la catégorie. Théoriquement, on peut compléter l'étude en traçant des ellipses de concentration ou, mieux, des enveloppes convexes des nuages de points [Junca 1985]

Dans la réalisation informatique, l'accent a été mis sur l'opérationnalité. Appelant en mémoire centrale les coordonnées factorielles de l'ensemble des individus, de même que leurs réponses sur l'ensemble des variables qualitatives, on a diminué au maximum les temps de calcul des répétitions bootstrap. A l'aide d'une souris, l'utilisateur peut, très rapidement, sélectionner les modalités à étudier ; en jouant sur les couleurs, il peut faire afficher à l'écran, simultanément, les nuages des points correspondant à différentes modalités d'une même variable (soit les coordonnées initiales des individus, soit les centres de gravité des répétitions bootstrap). De cette façon, il est possible d'obtenir une première idée, dans des délais raisonnables, de la stabilité de plusieurs centaines de points moyens. Des sorties imprimante ou table traçante sont, par ailleurs, possibles (voir figure 4.13).



*Enquête Agoramétrie 1987 : n=1470 individus*

On considère, ici, les coordonnées des centres de gravité des 50 échantillons bootstrap de deux modalités relatives à la variable "tendance politique" sur le plan principal 1-2 de la figure 4.5.

Le centre de gravité des "indifférents" (effectif : 209) paraît alors stable, tandis que les coordonnées (en particulier celle sur l'axe vertical) du centre de gravité des "extrême droite" (effectif : 14) sont assez instables.

FIGURE 4.13 : Dispersion de centres de gravité bootstrap

## BIBLIOGRAPHIE

- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*, 2nd edition. Wiley
- ten Berge, J. M. F. (1984). A joint treatment of varimax rotation and the problem of diagonalizing symmetric matrices simultaneously in the least squares sense. *Psychometrika*, 49, 3, 347-358.
- Cazes, P., S. Bonnefous, A. Baumerder & J. P. Pagès (1976). *Description cohérente des variables qualitatives prises globalement et de leur modalités*. *Statistique et Analyse des Données*, 2-3, 48-62.
- Golub, G. H. & C. Van Loan (1983). *Matrix Computations*. Baltimore : John Hopkins.
- Harman, H. H. (1968). *Modern Factor Analysis*. Chicago : University of Chicago Press.
- Hatabian, G. & G. Saporta (1985). *Regions de confiance en Analyse Factorielle*. Actes des 4èmes Journées Internationales "Informatique et Analyse des Données", INRIA, Versailles, 611-625.
- Hendrickson, A. E. & P. O. White. (1964). *PROMAX : a quick method for rotation in oblique simple structure*. *British Journal of Statistical Psychology*, 17, 65-70.
- Holmes-Junca, S. (1985). *Outils informatiques pour l'évaluation de la pertinence d'un résultat en Analyse des Données*. Thèse de troisième cycle. Université des Sciences et Techniques du Languedoc.
- Jennrich, R. I. & P. F. Sampson (1966). *Rotation for simple loadings*. *Psychometrika*, 31, 313-323.
- Joliffe, I. T. (1986). *Principal components analysis*. New-York : Springer Verlag.
- Laumon, B., J. L. Martin & D. Balland (1985). *Un algorithme de calcul de la probabilité qu'une variable de Fisher-Snédecour, à premier degré de liberté impair, excède une valeur donnée*. *Revue de la Statistique Appliquée*, XXXIII, 4, 29-37.
- Lebart, L. & A. Morineau (1985). *Système Portable pour l'Analyse des Données*. Paris : CESIA.
- Lebart, L., A. Morineau & K. M. Warwick (1984). *Multivariate descriptive statistical analysis*. Wiley
- Morineau, A. (1984). *Note sur la caractérisation statistique d'une classe et les valeurs-tests*. *Bulletin Technique du CESIA*, 2, 20-27.

- Saporta, G. (1979). *Pondération optimale de variables qualitatives en analyse de données*. *Statistique et Analyse des Données*, 3, 19-31.
- Sherin, R. J. (1966). *A matrix formulation to Kaiser's varimax criterion*. *Psychometrika*, 31, 535-538.
- Thionet, P. (1976). *Construction et reconstruction des tableaux statistiques*. *Annales de l'INSEE*, 22-23, 5-27, Paris.
- Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*. London : Oxford University Press.
- Zelen, M. & N.C. Severo (1964). *Probability functions*. Washington, IN : M. Abramowitz et al. ed., *Handbook of Mathematical Functions*, 925-964.

## CHAPITRE V

### EVALUATION DE LE DEMARCHE A L'AIDE D'EXEMPLES

#### INTRODUCTION

Dans ce chapitre, on illustre sur trois applications la méthodologie exposée dans le troisième chapitre. L'analyse d'enquêtes d'opinion particulières est effectuée à la lumière des structures factorielles dégagées lors des enquêtes nationales faites par Agoramétrie.

Avant d'aborder les trois exemples, on présente dans une première section le repère factoriel qu'Agoramétrie retient comme référence. Obtenu par analyse en composantes principales en Mars 1981, dans une période où la société française pouvait être considérée comme étant en équilibre, ce repère a été laborieusement interprété ; depuis, il sert de cadre d'analyse pour un grand nombre d'enquêtes particulières.

Le premier exemple concerne une pré-enquête sur l'action militante et la protection sociale, réalisée en 1987 par les étudiants de l'Ecole Nationale Supérieure des Pétroles et des Moteurs (IFP) en collaboration avec la société TEN. La présentation des résultats des analyses effectuées avec la technique de la greffe permet de souligner l'intérêt de telles pré-enquêtes réalisées à un coût extrêmement réduit, sans grande précaution au niveau de l'échantillonnage. Ayant précisé en quoi le discours sur le militantisme et la protection sociale relève des idéolo-

gies, nous montrons, en effet, pourquoi les résultats peuvent être généralisés à la population des Français.

Pour comparer les structures de l'opinion publique dégagées par Agoramétrie au cours du temps (diachronie), il est fait appel à la technique des rotations procrustéennes. On montre, dans le deuxième exemple, comment s'est effectué le recalage des dimensions factorielles de l'enquête de 1985 sur le repère de référence de Mars 1981.

Dans le troisième exemple, on aborde le problème de la perception des risques dans le grand public ; une enquête nationale a été réalisée en Février 1986 dans le cadre d'un groupe de travail inter-entreprises. S'étant replacé, par l'intermédiaire de l'enquête réalisée par Agoramétrie en 1985, dans l'enquête de référence de Mars 1981, on a pu mesurer l'importance des relations existant entre le risque "perçu" et les opinions sur les thèmes de conflit. L'analyse de redondance a permis de vérifier que ces indicateurs globaux d'opinion que sont les structures dégagées par Agoramétrie pouvaient être considérés comme des décodeurs, presque optimaux, pour interpréter les résultats d'une telle enquête.

### 5.1. Le repère de référence (enquête de Mars 1981)

Nous reprenons ici, pratiquement in extenso, l'interprétation des quatre premiers indicateurs de référence décrite dans [*Les structures de l'opinion en 1987*"], édité à Paris par l'association Agoramétrie]. Rappelons que ces indicateurs sont les dimensions principales extraites du tableau "individusxvariables" (1543x76) correspondant à l'enquête nationale réalisée par Agoramétrie en Mars 1981.

#### **Etiquetage des deux premières dimensions**

Pour interpréter les dimensions factorielles, il faut consulter simultanément les graphiques des figures 5.1.a (cercle des corélations) et 5.1.b (plan principal). Les axes des connotations dans l'espace des

conflits (composantes principales) et les axes de différenciation dans l'espace des individus (axes principaux) étant en parfaite dualité (cf section 2.2.), les quatre quadrants numérotés 1,2,3 et 4 dans les deux graphiques se correspondent ; en les associant : 1 avec 1, 2 avec 2, ... on relie les groupes sociaux aux opinions sur les conflits qui les caractérisent.

Dans le premier quadrant où se retrouvent les moins diplômés et les plus âgés on est comme dans une forteresse : on y défend les valeurs, les règles, les normes ou les stéréotypes dont s'est dotée la société pour résoudre ses problèmes ; et en même temps, parfois de façon paradoxale, on y dénonce les dysfonctionnements sociaux. Ce premier quadrant du repli s'oppose au troisième qui est celui des grandes offensives et de l'aventure. C'est là que les plus jeunes, les mieux fournis en diplômes et ceux que l'on appelle parfois les intellectuels de gauche développent leurs attaques contre les interdits ; c'est là qu'en 1981 se fabriquaient véritablement les idées nouvelles. La polémique qui oppose les deux autres quadrants ne porte pas, elle, sur les règles et les valeurs. Les propositions qui sont faites par ceux qui peuplent le deuxième quadrant, et qui en majorité se reconnaissent en 1981 dans l'opposition de gauche, constituent un véritable défi ; au-delà des contradictions, on en demande plus en matière de liberté, d'égalité et de solidarité tout à la fois. Le défi vient du fait que la revendication paraît légitime : on en revient aux prémices, à la déclaration des droits de l'homme qui apparaît comme le fondement de l'Etat républicain. Mais une certaine distance ne sépare-t-elle pas toujours la théorie de la pratique, l'idéal de la réalité, l'intention de l'action ? Les sociétés résultent d'un apprentissage ; compromis fournis par l'histoire, elles sont soumises en permanence à l'épreuve des faits. Aussi ne serait-ce pas plutôt à l'idée de "la grande marche"<sup>1</sup> à laquelle on adhérerait en soutenant ce vaste programme qui peut être renouvelé à l'infini ? Ce qui est certain, c'est que ceux du quatrième quadrant, qui se situaient en majorité au centre ou à droite en mars 1981, semblent insensibles à cette idée superbe ! A l'idéal déstabilisateur, ils opposent le pragmatisme et le réalisme : en proposant des mesures qui ne peuvent compromettre les équilibres, ils adoptent une attitude de soumission à la réalité.

---

<sup>1</sup> Kundera, Milan (1984). *L'insoutenable légèreté de l'être*. Gallimard.

FIGURE 5.1 : Les dimensions de référence d'Agoramétrie (Suite...)

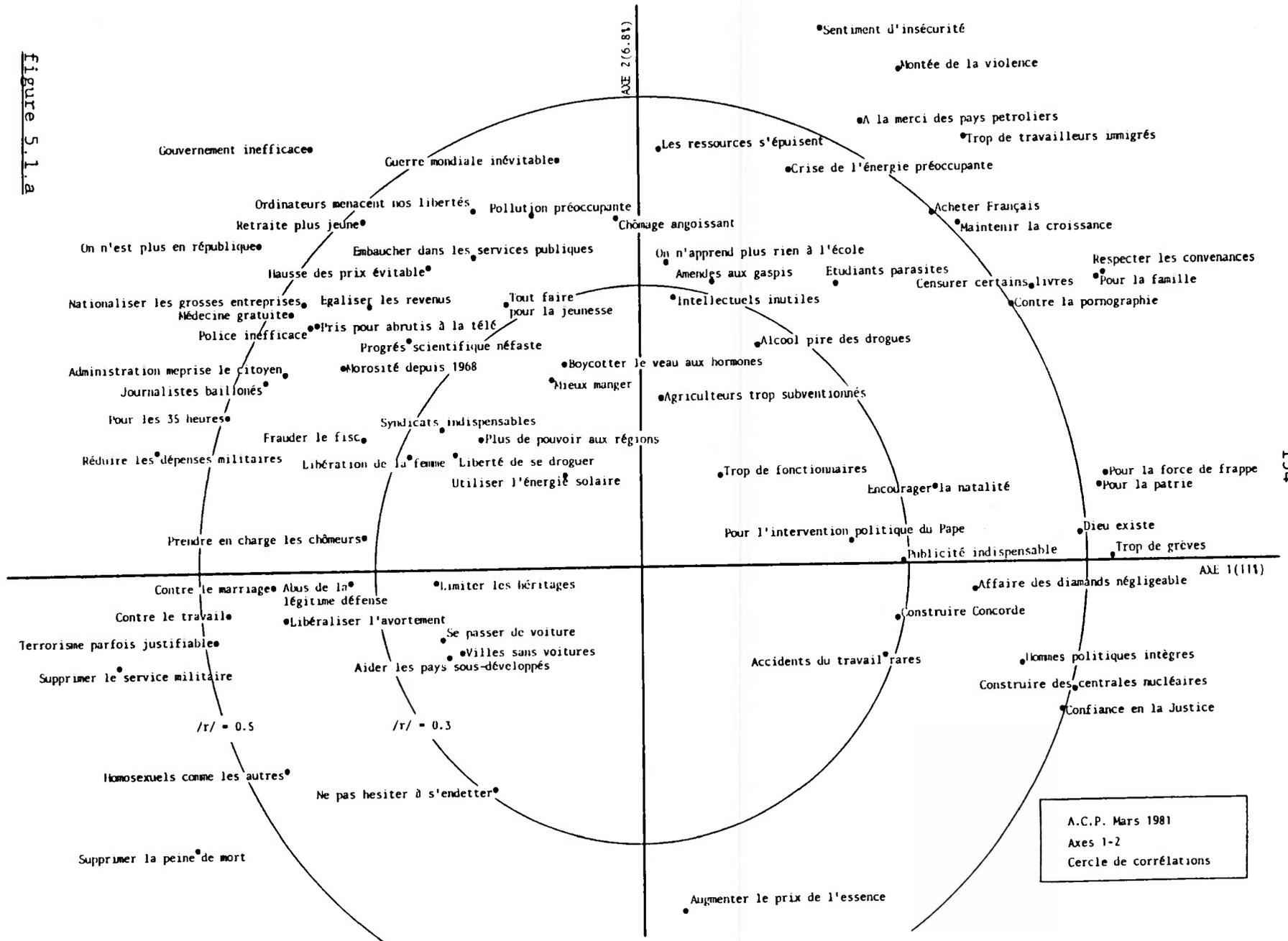


Figure 5.1.b

FIGURE 5.1 : (...Suite...)

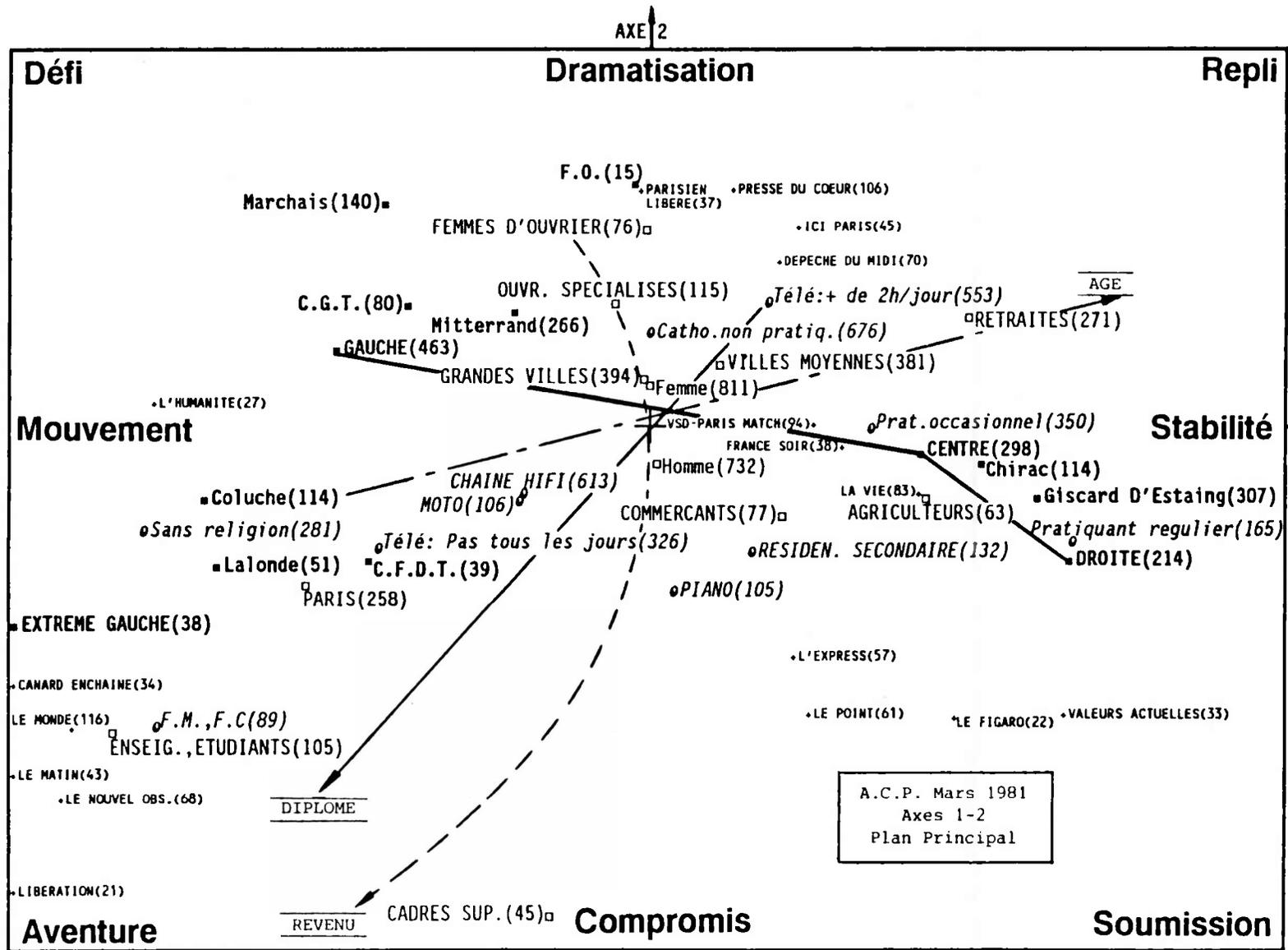
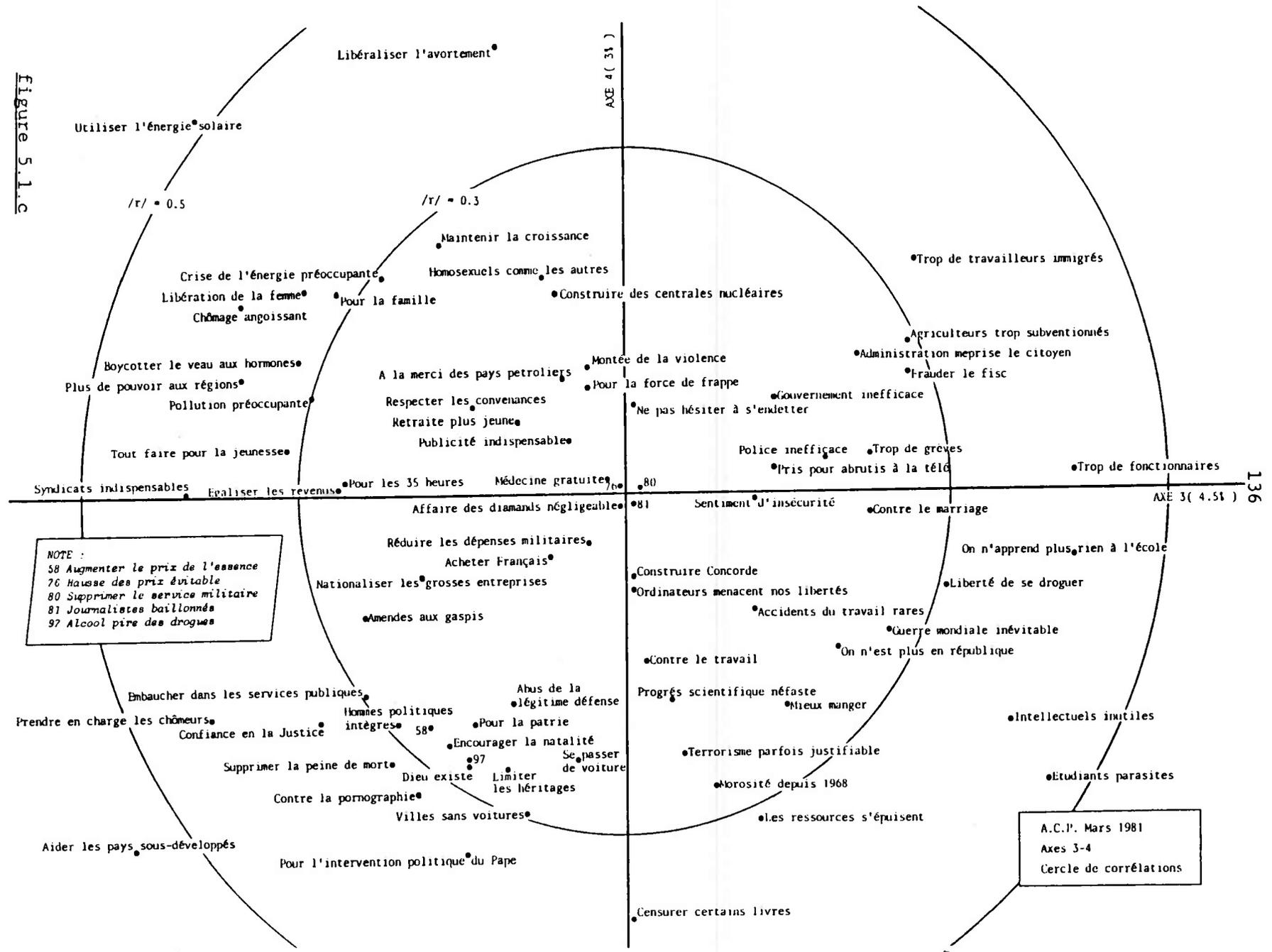


Figure 5.1.c

FIGURE 5.1 : (...Suite...)

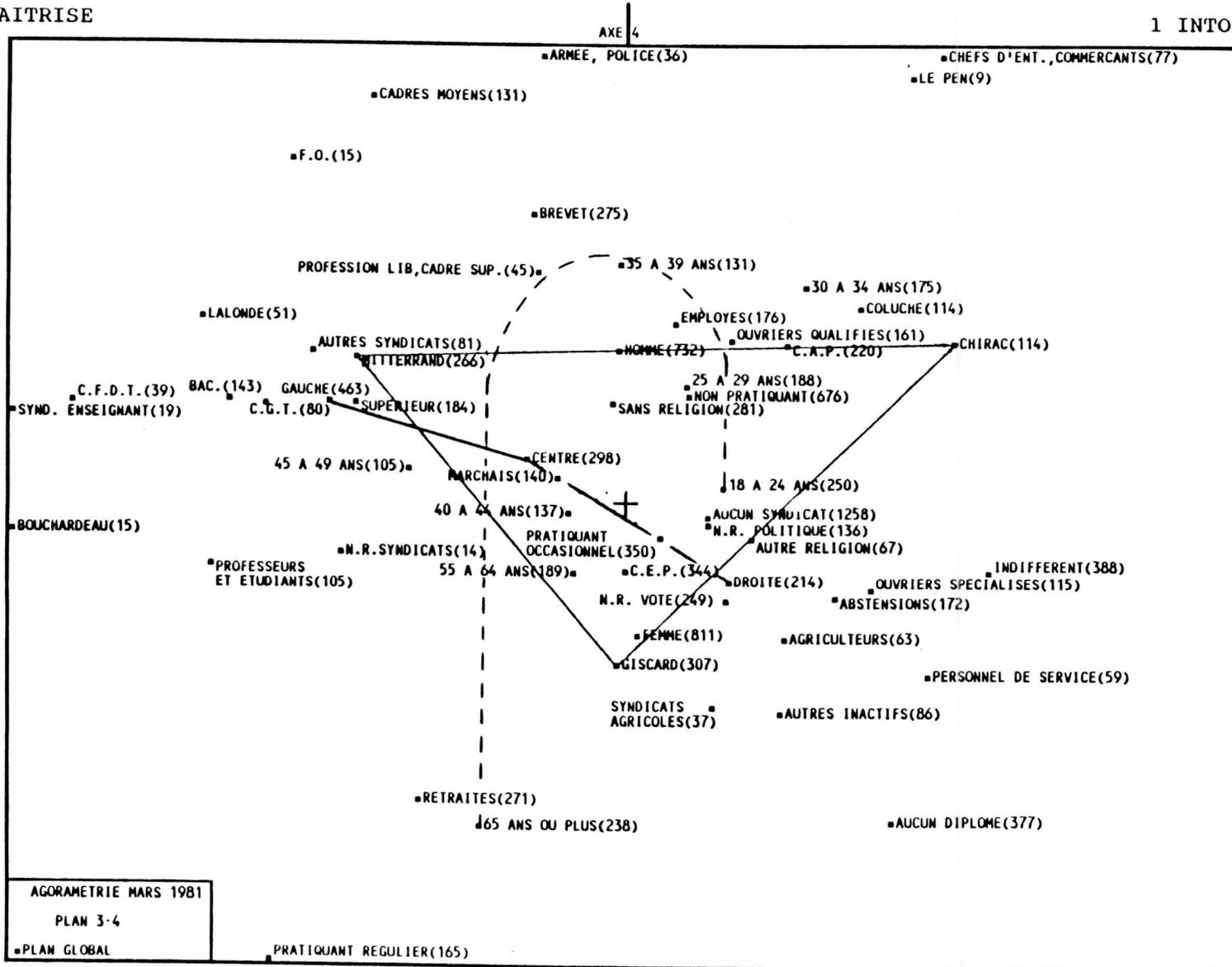


2 MAITRISE

INDIVIDUEL

1 INTOLERANCE

Figure 5.1.d



AXE 3

MOROSITE

137

BEATITUDE

FIGURE 5.1 : (...Suite)

AGORAMETRIE MARS 1981

PLAN 3-4

PLAN GLOBAL

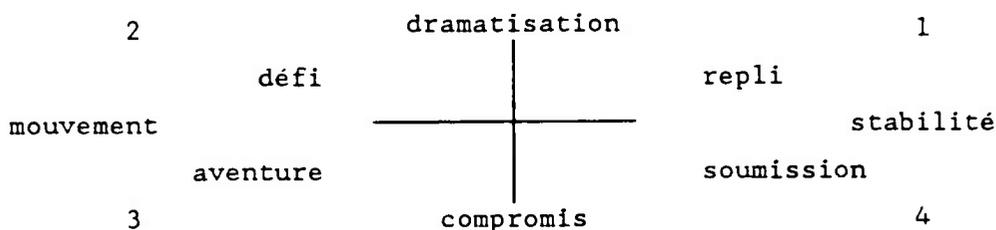
3 TOLERANCE

COLLECTIF

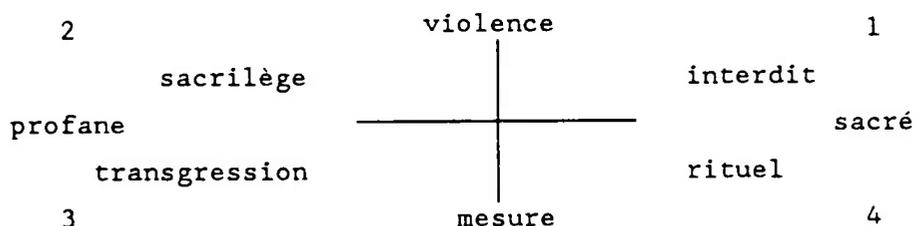
4 IMPUISSANCE

Ces quatre catégories d'attitudes qui correspondent aux quatre quadrants apparaissent comme des combinaisons des attitudes élémentaires qui semblent caractériser les deux dimensions factorielles : la première dimension "de la différenciation" serait celle qui oppose la stabilité au mouvement ; la seconde opposerait la dramatisation au compromis.

Ces premières impressions fournissent donc un jeu de huit étiquettes qui permet de baliser le premier plan factoriel.



Ces étiquettes, dont une véritable algèbre assure la cohérence, n'ont d'intérêt que dans la mesure où elles inspirent l'utilisateur ; si ce dernier veut suivre René Girard<sup>2</sup>, il préférera certes - c'est son droit - un autre balisage.



Ici la violence et le sacré sont devenus les dimensions principales qui déterminent l'ensemble des attitudes : l'interdit et le rituel sont respectivement la forme violente et la forme mesurée du sacré ; ces formes s'opposent au sacrilège et à la transgression qui sont, de même, l'expression violente et l'expression mesurée du profane. On voit l'intérêt - et le danger ! - de ces nouvelles étiquettes qui suggèrent de puiser dans René Girard pour expliquer ce qui se passera au niveau des opinions.

---

<sup>2</sup> Girard, R. (1972). *La violence et le sacré*. Paris : Grasset.

On peut ainsi associer bien des jeux d'étiquettes, qui ne sont pas en contradiction les uns avec les autres, aux deux premières dimensions factorielles ; pour le systémicien, tous ces jeux sont sensiblement équivalents. On tend, plus qu'ailleurs :

- . à figer le système dans le premier quadrant,
- . à le déstabiliser dans le second,
- . à le faire transiter vers un nouvel état d'équilibre dans le troisième,
- . à préserver l'équilibre dans lequel il se trouve dans le quatrième quadrant.

### Les dimensions trois et quatre

De même que pour les deux premières dimensions, il faut consulter simultanément le plan principal (figure 5.1.c) et le cercle des corrélations (figure 5.1.d) pour interpréter les dimensions trois et quatre.

Regardons le cercle des corrélations : c'est bien la morosité, l'amertume et peut-être même la grogne qui semble s'exprimer le long de l'axe 3 : d'un côté, les attitudes sont plus négatives vis-à-vis du fisc, des travailleurs immigrés, des agriculteurs, de l'école, des fonctionnaires, des étudiants et des intellectuels... tout à la fois ; de l'autre, c'est tout le contraire, on soutient plus qu'ailleurs indifféremment : le solaire, la jeunesse, la femme, les syndicats, les chômeurs et les pays sous-développés. Est-ce la béatitude ? Le caractère systématiquement négatif (ou positif) des positions prises suggère ces interprétations.

Et maintenant regardons le plan principal : au hit parade de la morosité se distinguaient en mars 1981 : les commerçants, les ouvriers spécialisés, les agriculteurs et les "autres inactifs", les "sans diplôme", les revenus plus que modestes et ceux qui avaient manifesté, deux mois avant l'élection, l'intention de s'abstenir ou de voter Chirac ou Coluche au premier tour des présidentielles (celles de 1981). De l'autre côté, à la palme de la béatitude la partie était serrée entre les bacheliers et diplômés supérieurs, les professeurs et étudiants, les

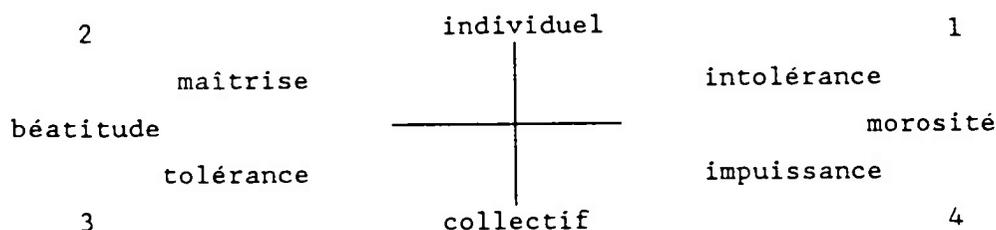
cadres moyens, ceux qui avaient des sympathies pour les syndicats ou pour la gauche... et les pratiquants réguliers.

En faisant éclater en quatre catégories d'attitudes antagonistes deux à deux l'opposition morosité - béatitude, la quatrième dimension permet de mettre de l'ordre dans cette cacophonie :

. la grogne "poujadiste" qui s'exprime dans le premier quadrant prend pour cibles privilégiées l'administration, le fisc et les immigrants ; son contraire est cette béatitude quasi céleste qu'évoquent les attitudes prises dans le troisième quadrant où l'on semble s'être donné le mot pour faire confiance aux institutions (au Pape en particulier) et aider son prochain. A l'intolérance affichée par le marchand (les commerçants) s'oppose donc ici la tolérance prônée par le croyant (les catholiques pratiquants) ;

. la confrontation entre les deux autres quadrants se développe, le long de la seconde bissectrice, sur un tout autre registre : la connaissance et le savoir sont ici les enjeux. Les thèmes ayant cours dans l'intelligentsia (thèmes écologiques ou de "libération") ont manifestement beaucoup plus de succès dans le deuxième quadrant, où se retrouvent les plus diplômés et à un degré moindre les plus riches, que dans le quatrième où la grogne des plus démunis ("autres inactifs" et "sans diplôme") se déchaîne avant tout contre les étudiants et les intellectuels. Par le rejet de l'élite et de la connaissance, voilà donc comment s'exprime l'impuissance face à la maitrise du pouvoir et du savoir.

Ainsi la quatrième dimension apporte un éclairage sur le jeu de la morosité et de la béatitude. Marquerait-elle une opposition entre le collectif et l'individuel ? Sans se prononcer définitivement, on peut l'admettre au nom de la cohérence algébrique que doit respecter l'ensemble des étiquettes.



## 5.2. Un éclairage sur l'action militante et la protection sociale

Le repère de référence pour la greffe est ici formé des trois premières composantes principales issues de l'ACP effectuée sur l'enquête Agoramétrie 1987. Si on consulte le cercle des corrélations 1-2 et le plan principal donnés dans la figure 5.2, les dimensions factorielles obtenues en 1987 paraissent proches de leurs homologues obtenues en Mars 1981 (comparer les figures 5.1.a et 5.2.a). Les corrélations entre composantes principales du même ordre, telles qu'elles ont été calculées par la technique des rotations procrustéennes, sont élevées ; dans l'ordre, pour les trois premières composantes : 0.900, 0.789, 0.916. Aussi, on ne saurait associer à ces dimensions de 1987 d'autres étiquettes que celles affectées aux axes extraits en Mars 1981.

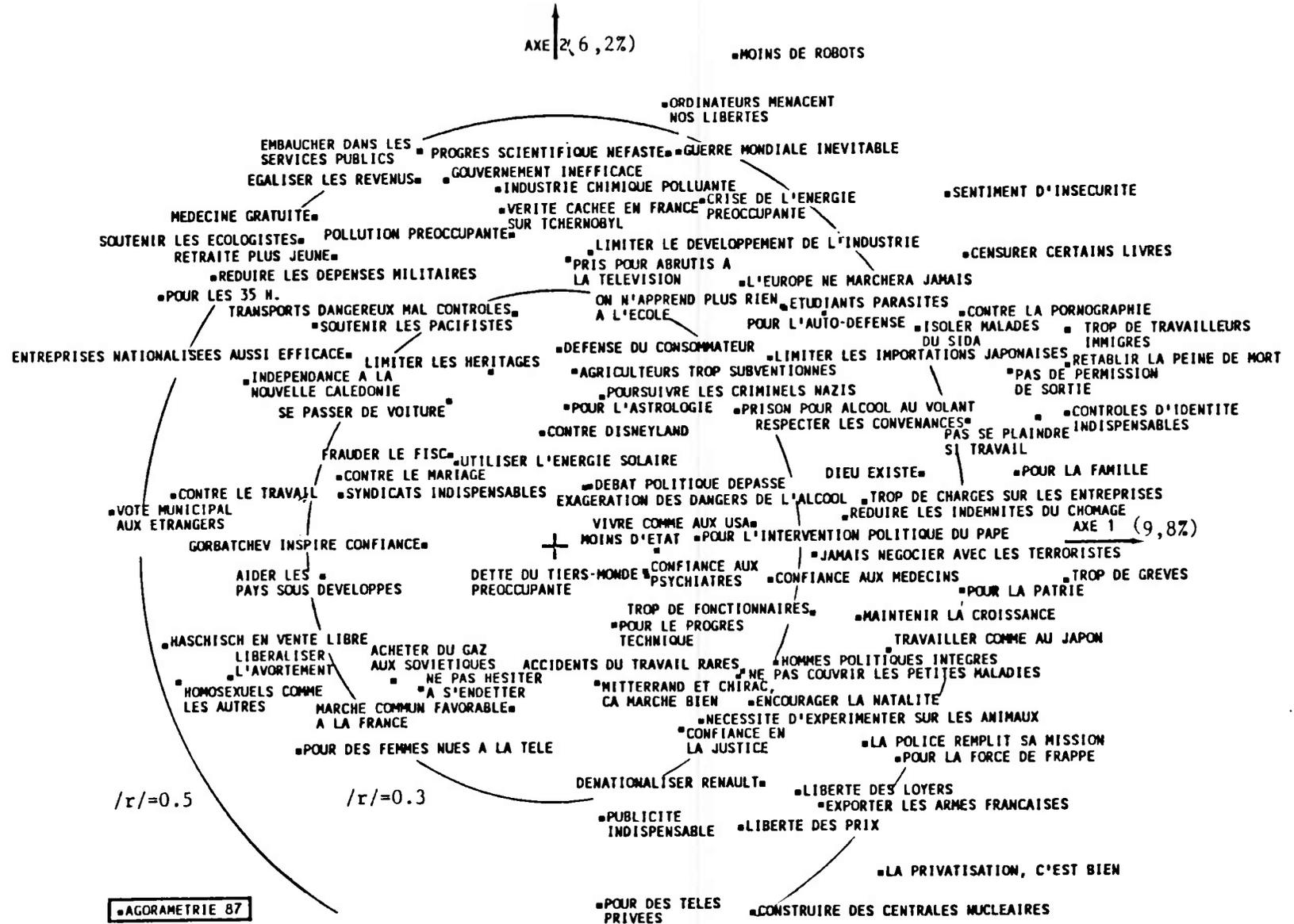
### Questionnaire - Composition de l'échantillon

Construit dans le but de préparer une enquête plus importante sur le militantisme et la protection sociale, le questionnaire résulte d'un "brainstorming". Il comprend d'abord sept parties traitant successivement des sujets qui préoccupent le plus les Français (voir figure 5.5), des grandes causes auxquelles on accepterait de se consacrer, de l'image de certaines vedettes de la politique et des médias, de l'image des organisations oeuvrant dans le champ de la protection sociale, du problème du chômage, du problème plus général de la protection sociale et, enfin, des droits acquis que l'on est prêt à sacrifier pour sortir de la crise. Une huitième partie est constituée uniquement de propositions (thèmes de conflit) tirées de l'enquête "Agoramétrie 1987" ; elles permettront le recours aux indicateurs globaux d'opinion afin d'éclairer les positions prises sur des questions qui relèvent uniquement du social. La dernière partie du questionnaire constitue le "signalétique" ; il comprend un petit nombre de variables socio-démographiques et socio-culturelles : âge, diplôme, sexe, catégorie socio-professionnelle, religion, et tendance politique.

Les étudiants de l'IFP ont enquêté dans leur entourage. Aussi, la composition de l'échantillon des 247 personnes interrogées est-elle très différente de celle de l'enquête nationale de référence où l'on a affaire à un échantillon représentatif de la population des Français

FIGURE 5.2 : Référent d'Agoramétrie 1987 (Suite...)

Figure 5.2.a : cercle des corrélations





6. LES POINTS SENSIBLES						7. LES THEMES D'INTERET							
<p>Jugez-vous les sujets suivants</p> <p>1 = Pas du tout préoccupants  2 = Pas tellement préoccupants  3 = Préoccupants  4 = Très préoccupants  5 = Extrêmement préoccupants</p>						<p>Accepteriez-vous de consacrer du temps et de l'argent pour les causes suivantes</p> <p>1 = Pas du tout d'accord  2 = Pas tellement d'accord  3 = Peut-être d'accord  4 = Bien d'accord  5 = Tout à fait d'accord</p>							
		1	2	3	4	5			1	2	3	4	5
1	Déficit Sécurité Sociale						1	Recherche médicale					
2	Chômage						2	Aide aux personnes âgées					
3	Immigrés						3	Lutte contre le racisme					
4	Crise du logement						4	Défense des consommateurs					
5	Toxicomanie						5	Amélioration des conditions du travail					
6	Sida						6	Faim dans le monde					
7	Financement des retraites						7	Lutte pour la paix					
8	Pollution						8	Lutte contre le SIDA					
9	Insécurité						9	Défense de l'environnement					
10	Pauvreté						10	Egalité des sexes					
11	Accidents de la route						11	Développement du sport					
12	Personnes âgées						12	Aide aux pauvres					
13	Alcoolisme						13	Réinsertion des prisonniers					
14	Recul du niveau de vie						14	Défense des valeurs traditionnelles					
15	Terrorisme						15	Aide aux handicapés					
16	Remise en cause des valeurs						16	Défense du pouvoir d'achat					
17	Risque industriel						17	Aide aux pays sous-développés					
18	Prostitution						18	Défense de l'emploi					
19	Handicapés						19	Animation des quartiers					
20	Délinquance des jeunes						20	Défense de la propriété					
21	Baisse de la natalité						21	Lutte contre l'insécurité					
22	Cohabitation politique						22	Action évangélique					
23							23	Développement de la vie associative					
24							24	Lutte contre la prostitution					
25							25	Lutte contre les sectes					
26							26	Secourisme					
27							27	Aide aux drogués					
28							28	Action syndicale					
29							29	Amélioration de l'enseignement					
30							30	Liberté d'expression					

FIGURE 5.3 : Extrait du questionnaire

majeurs vivant en métropole. Les principales divergences sont résumées dans le tableau suivant :

		AGORAMETRIE n=1470	PRE-ENQUETE n=247
AGE	18 à 24 ans	15.3%	41.3%
	25 à 34 ans	24.4%	30.4%
	35 à 49 ans	23.9%	11.7%
	50 à 64 ans	21.0%	15.8%
	65 ans et +	15.3%	0.8%
CSP	Cad. sup., Prof. libérales	4.8%	19.4%
	Enseignants, Etudiants	12.2%	41.3%
	Technicien, Cadre moyen	7.3%	11.7%
	Employés	16.1%	12.1%
	Autres Csp	61.8%	15.4%
REGION	Region Parisienne	18.1%	100.0%

### Construction de la passerelle

Dans l'enquête "Agoramétrie 1987" (enquête A) l'ACP a porté sur 96 variables. Les variables reconstruisant au mieux les quatre premières composantes principales, au sens des coefficients de corrélation multiple, ont été extraites avec la deuxième des procédures décrites dans le paragraphe 3.2.2.. Les cinq premières variables "optimales", accompagnées de dix-neuf autres jugées intéressantes pour l'étude menée, forment la partie "thèmes de conflit" du questionnaire de la pré-enquête (enquête B). Avec les libellés réduits des vingt-quatre variables, apparaissent dans le tableau de la figure 5.4.a les coefficients de régression qu'il faut connaître pour effectuer la greffe. On remarquera la valeur élevée du coefficient de corrélation multiple associé à la reconstruction de la première composante ; les autres composantes, en particulier la quatrième, sont nettement moins bien reconstruites.

	C1	C2	C3	C4	MOY	ECT
valeurs propres :	9.429	5.949	4.129	3.133		
LIBERALISER L'AVORTEMENT	-0.071	-0.061	-0.032	0.225	3.380	1.481
CONSTRUIRE DES CENTRALES NUCLEAIRES	0.071	-0.239	0.002	0.126	2.295	1.244
DIEU EXISTE	0.095	0.034	-0.098	-0.049	3.338	1.439
CONFIANCE EN LA JUSTICE	0.060	-0.103	-0.329	0.208	2.743	1.153
TROP DE TRAVAILLEURS IMMIGRES	0.176	0.125	0.160	0.183	3.404	1.435
CONTRE LA PORNOGRAPHIE	0.107	0.123	-0.151	-0.111	3.197	1.452
AIDER LES PAYS SOUS DEVELOPPES	-0.052	-0.022	-0.173	0.092	3.198	1.296
POLLUTION PREOCCUPANTE	-0.023	0.167	-0.124	-0.114	4.389	0.855
POUR LA PATRIE	0.115	-0.019	-0.151	0.189	2.525	1.273
ENCOURAGER LA NATALITE	0.038	-0.058	-0.122	0.091	3.299	1.235
REDUIRE LES DEPENSES MILITAIRES	-0.102	0.173	0.011	0.012	3.618	1.273
SENTIMENT D'INSECURITE	0.109	0.184	0.004	0.020	3.595	1.325
POUR LES 35 H.	-0.104	0.155	-0.019	0.193	3.037	1.467
EMBAUCHER DANS LES SERVICES PUBLICS	-0.041	0.223	-0.186	0.131	3.361	1.351
RETRAITE PLUS JEUNE	-0.055	0.128	0.076	0.188	3.093	1.417
TROP DE GREVES	0.148	0.014	0.090	0.016	3.210	1.421
RETABLIR LA PEINE DE MORT	0.161	0.100	0.093	0.107	3.252	1.636
CONTROLES D'IDENTITE INDISPENSABLES	0.167	0.020	-0.163	0.034	3.705	1.351
REDUIRE LES INDEMNITES DU CHOMAGE	0.071	0.011	0.134	0.131	2.111	1.201
POUR DES TELES PRIVEES	0.011	-0.177	0.037	0.008	3.340	1.320
TROP DE CHARGES SUR LES ENTREPRISES	0.071	-0.002	0.086	-0.034	3.754	1.214
INDEPENDANCE A LA NOUVELLE CALEDONIE	-0.078	0.087	-0.040	0.020	3.264	1.187
POUR DES FEMMES NUES A LA TELE	-0.051	-0.075	0.151	0.229	2.759	1.342
GORBATCHEV INSPIRE CONFIANCE	-0.017	-0.027	-0.177	0.164	2.669	1.098
COEFFICIENTS DE CORRELATION MULTIPLE :	0.952	0.847	0.814	0.745		

figure 5.4.a. Coefficients de régression permettant de reconstruire les 4 premières composantes principales (normées) et moyenne et écart-type des différentes variables dans l'enquête Agoramétrie

COMPOSANTE N°1	MOYENNE	EC. TYPE	PROBA du T
LIBERALISER L'AVORTEMENT	-0.071	0.009	1.12E-013
CONSTRUIRE DES CENTRALES NUCLEAIRES	0.071	0.009	6.22E-015
DIEU EXISTE	0.095	0.009	9.99E-016
CONFIANCE EN LA JUSTICE	0.060	0.009	2.85E-011
TROP DE TRAVAILLEURS IMMIGRES	0.176	0.011	1.00E-016
CONTRE LA PORNOGRAPHIE	0.107	0.011	5.55E-016
AIDER LES PAYS SOUS DEVELOPPES	-0.052	0.009	3.10E-008
POLLUTION PREOCCUPANTE	-0.023	0.009	7.89E-003
POUR LA PATRIE	0.115	0.010	4.44E-016
ENCOURAGER LA NATALITE	0.038	0.009	2.65E-005
REDUIRE LES DEPENSES MILITAIRES	-0.102	0.009	1.00E-016
SENTIMENT D'INSECURITE	0.109	0.010	1.78E-015
POUR LES 35 H.	-0.104	0.010	1.55E-015
EMBAUCHER DANS LES SERVICES PUBLICS	-0.041	0.009	7.72E-006
RETRAITE PLUS JEUNE	-0.055	0.009	6.44E-009
TROP DE GREVES	0.148	0.010	1.00E-016
RETABLIR LA PEINE DE MORT	0.161	0.011	1.11E-016
CONTROLES D'IDENTITE INDISPENSABLES	0.167	0.010	1.00E-016
REDUIRE LES INDEMNITES DU CHOMAGE	0.071	0.009	4.22E-015
POUR DES TELES PRIVEES	0.011	0.009	2.35E-001
TROP DE CHARGES SUR LES ENTREPRISES	0.071	0.009	9.66E-015
INDEPENDANCE A LA NOUVELLE CALEDONIE	-0.078	0.009	3.11E-015
POUR DES FEMMES NUES A LA TELE	-0.051	0.010	4.43E-007
GORBATCHEV INSPIRE CONFIANCE	-0.017	0.009	5.33E-002

figure 5.4.b. Résultats des tests t (probabilités de dépassement) effectués sur chacun des coefficients de régression pour la première composante

FIGURE 5.4 : Coefficients de greffe

Les très nettes différences dans la composition des échantillons se retrouvent dans les pourcentages : 70% des interviewés dans l'enquête Agoramétrie déclarent ressentir de l'insécurité, contre 4.9% seulement dans la pré-enquête. Mais ces différences semblent affecter beaucoup moins les corrélations. En consultant le tableau comparatif de la figure 5.5.a, on constate en effet que, si on note quelques écarts importants (de l'ordre de 0.20) entre coefficients de corrélation homologues, les corrélations vont, néanmoins, dans le même sens dans les deux enquêtes (il n'y a pas d'inversion de signes). Les plus fortes d'entre elles, par ailleurs, sont assez stables, comme on peut en juger en examinant, par exemple, les coefficients de corrélation supérieurs à 0.30 de la variable "sentiment d'insécurité".

Théoriquement, on devrait écarter de la passerelle toutes les variables ayant dans les deux enquêtes des profils de corrélation éloignés. En utilisant la procédure d'élimination décrite dans le paragraphe 3.2.3, il aurait fallu écarter la bonne moitié des variables communes pour aboutir aux douze variables du tableau de la figure 5.5.b. Cependant, ici les 24 variables communes ont été finalement retenues dans la passerelle et cela pour différentes raisons. En effet :

- nous n'avons pas été capable de donner une justification psychosociale à l'élimination d'une quelconque des variables.

- les matrices de corrélation ne devenant voisines qu'après élimination de plusieurs variables, nous avons constaté que la qualité des reconstructions des indicateurs factoriels dans A était insuffisante si on se restreignait à douze variables.

- les études de robustesse effectuées ont montré une remarquable stabilité des résultats : les différentes régressions effectuées ayant écarté les variables les plus "éloignées" et les différents essais de recalage par rotations procrustéennes (en écartant aussi de la cible les variables les plus "éloignées") ont fourni, en gros, les mêmes résultats.

Donnons, à titre d'exemple, les corrélations dans l'enquête B entre composantes  $\hat{c}^l$  obtenues par régression sur 24 variables et composantes  $r^l$  obtenues par rotation des 6 premières composantes principales issues de l'ACP dans B sur 24 variables (matrice cible A(16x3) constituée des coordonnées sur les trois premières composantes de l'enquête A des 16 variables les moins "éloignées" ; coefficient de "dilatation"  $d=1.13$ ).

24 VARIABLES

	LIBERALISER L'AVORTEMENT	CONSTRUIRE DES CENTRALES NUCLEAIRES	DIEU EXISTE	CONFIANCE EN LA JUSTICE	TROP DE TRAVAILLEURS IMMIGRES	CONTRE LA PORNOGRAPHIE	AIDER LES PAYS SOUS DEVELOPPES	POLLUTION PREOCCUPANTE	POUR LA PATRIE	ENCOURAGER LA NATALITE	REDUIRE LES DEPENSES MILITAIRES	SENTIMENT D'INSECURITE	POUR LES 35 H.	EMBAUCHER DANS LES SERVICES PUBLICS	RETRAITE PLUS JEUNE	TROP DE GREVES	RETABLIR LA PEINE DE MORT	CONTROLES D'IDENTITE INDISPENSABLES	REDUIRE LES INDEMNITES DU CHOMAGE	POUR DES TELES PRIVEES	TROP DE CHARGES SUR LES ENTREPRISES	INDEPENDANCE A LA NOUVELLE CALEDONIE	POUR DES FEMMES NUES A LA TELE	GORBATCHEV INSPIRE CONFIANCE
LIBERALISER L'AVORTEMENT	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
CONSTRUIRE DES CENTRALES NUCLEAIRES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
DIEU EXISTE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
CONFIANCE EN LA JUSTICE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
TROP DE TRAVAILLEURS IMMIGRES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
CONTRE LA PORNOGRAPHIE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
AIDER LES PAYS SOUS DEVELOPPES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
POLLUTION PREOCCUPANTE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
POUR LA PATRIE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ENCOURAGER LA NATALITE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
REDUIRE LES DEPENSES MILITAIRES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
SENTIMENT D'INSECURITE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
POUR LES 35 H.	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
EMBAUCHER DANS LES SERVICES PUBLICS	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
RETRAITE PLUS JEUNE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
TROP DE GREVES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
RETABLIR LA PEINE DE MORT	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
CONTROLES D'IDENTITE INDISPENSABLES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
REDUIRE LES INDEMNITES DU CHOMAGE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
POUR DES TELES PRIVEES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
TROP DE CHARGES SUR LES ENTREPRISES	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
INDEPENDANCE A LA NOUVELLE CALEDONIE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
POUR DES FEMMES NUES A LA TELE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
GORBATCHEV INSPIRE CONFIANCE	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

figure 5.5.a (pour les explications voir figure 5.5.b)

FIGURE 5.5 : Comparaison des matrices de corrélation entre variables communes (Suite...)

12 VARIABLES		C	D	C	C	A	E	S	R	P	I	P	G
Reinsertion de CONFIANCE EN LA JUSTICE a la place de POUR LES 35 H.		CON	DIEU	CON	CON	AIDER	ENCOUR	SENTI	RETA	POUR	INDEP	POUR	GOR
		STRUI	EXIS	FIAN	TRRE	LES	OURA	MENT	BLIR	DES	PEND	DES	BAT
		RE	TE	CE	LA	LES	GER	LA	LA	TELES	DANCE	FEM	TCHE
		DES	EN	LA	POR	PAYS	LA	D'	PEI	PRIVE	A	MES	V
		CEN	JUS	JUS	NOG	S	NAT	INSE	NE	NO	LA	NUE	INS
		TR	T	T	RAPH	S	ALI	CU	DE	NO	NUE	S	PIRE
					H	S	I	R					C
CONSTRUIRE DES CENTRALES NUCLEAIRES		4	14	-4	-5	11	0	13	20	-15	9	2	
		10	24	4	-10	21	-1	20	25	-17	10	0	
DIEU EXISTE		4	10	36	3	17	19	17	4	-13	-25	-11	
		10	14	47	9	24	14	15	2	-24	-33	-15	
CONFIANCE EN LA JUSTICE		14	10		12	14	21	-5	0	9	-9	-5	7
		24	14		13	7	22	0	-3	17	-23	-3	-1
CONTRE LA PORNOGRAPHIE		-4	36	12		3	11	28	19	-11	-9	-52	-3
		4	47	13		11	19	34	19	-13	-15	-53	-12
AIDER LES PAYS SOUS DEVELOPPES		-5	3	14	3		8	-16	-33	0	15	-5	6
		-10	9	7	11		0	-9	-23	-8	23	-10	11
ENCOURAGER LA NATALITE		11	17	21	11	8		4	7	11	-11	-5	-3
		21	24	22	19	0		12	6	20	-7	-13	-10
SENTIMENT D'INSECURITE		0	19	-5	28	-16	4		40	-12	-11	-18	-7
		-1	14	0	34	-9	12		42	-10	-11	-20	1
RETABLIR LA PEINE DE MORT		13	17	0	19	-33	7	40		-3	-18	-9	-6
		20	15	-3	19	-23	6	42		8	-18	0	-6
POUR DES TELES PRIVEES		20	4	9	-11	0	11	-12	-3		-7	20	-3
		25	2	17	-13	-8	20	-10	8		-12	28	-8
INDEPENDANCE A LA NOUVELLE CALEDONIE		-15	-13	-9	-9	15	-11	-11	-18	-7		5	16
		-17	-24	-23	-15	23	-7	-11	-18	-12		11	24
POUR DES FEMMES NUES A LA TELE		9	-25	-5	-52	-5	-5	-18	-9	20	5		4
		10	-33	-3	-53	-10	-13	-20	0	28	11		15
GORBATCHEV INSPIRE CONFIANCE		2	-11	7	-3	6	-3	-7	-6	-3	16		4
		0	-15	-1	-12	11	-10	1	-6	-8	24		15

figure 5.5.b

Enquêtes : A : Agoramétrie 1987 (1470 individus)  
 B : Protection Sociale (247 individus)  
 nombre de variables communes q=24.

Pour chaque variable, on donne dans la première ligne les corrélations dans A et dans la deuxième les corrélations dans B. Les coefficients de corrélation sont multipliés par cent et arrondis. Les différences supérieures à 0.10 sont encadrées, celles supérieures à 0.15 apparaissent sur fond grisé.

Dans la procédure d'élimination pas à pas descendante, la variable sortante est rayée

FIGURE 5.5 : (...Suite)

$r^{\ell} \hat{c}^{\ell}$	$\ell=1$	2	3
$\ell=1$	0.985	-0.140	0.025
2	-0.050	0.956	-0.100
3	0.006	-0.101	0.936

Corrélations entre composantes reconstruites par régression ( $\hat{c}^{\ell}$ ) et par rotation ( $r^{\ell}$ )

#### Qualité de la greffe (reconstructions par régression)

Dans la formule de reconstruction donnant les approximations  $\hat{c}^{\ell}$  des composantes factorielles  $c^{\ell}$  de A :

$$\hat{c}_i^{\ell} = \sum_{j=1}^q b_j^{\ell} \frac{z_i^j - m_j}{s_j} ; i=1, \dots, 247 \quad (3.13)$$

les  $q=24$  variables communes  $z^j$  sont donc intervenues. Les coefficients de régression  $b_j^{\ell}$ ,  $\ell=1, \dots, 4$ , la moyenne  $m_j$  et l'écart-type  $s_j$  des  $z^j$  dans A apparaissent dans le tableau de la figure 5.4.a. Les tests  $t$  effectués sont tous significatifs (sauf, éventuellement, pour certains coefficients très petits en valeur absolue). On peut s'en rendre compte en ce qui concerne la première composante en consultant le tableau de la figure 5.4.b.

La première vérification à effectuer dans B concerne l'orthogonalité entre les composantes  $\hat{c}^{\ell}$ . Le tableau suivant montre que les composantes  $c^{\ell}$ , telles qu'elles sont reconstruites, peuvent être considérées comme orthogonales, compte tenu de la précision exigée, malgré le -0.210 entre les composantes deux et trois<sup>1</sup>.

$\hat{c}^{\ell} \hat{c}^{\ell}$	$\ell=1$	2	3	4
$\ell=1$	1.000	-0.183	0.038	-0.009
2	-0.183	1.000	-0.210	-0.043
3	0.038	-0.210	1.000	-0.071
4	-0.009	-0.043	-0.071	1.000

Corrélations entre composantes  $\hat{c}^{\ell}$  obtenues par régression

<sup>1</sup>Soulignons que l'orthogonalité entre les  $\hat{c}^{\ell}$  ne témoigne pas, à elle seule, de la qualité de la greffe. Les variables d'opinion entretiennent souvent des corrélations relativement faibles. Deux combinaisons linéaires aléatoires des variables ont donc peu de chances d'être bien corrélées entre elles. Dans un essai de chaotisation, chaque composante  $\hat{c}_i^{\ell}$  a été reconstruite plusieurs fois en mélangeant les coefficients  $b_j^{\ell}$  de façon aléatoire : environ 30% des configurations obtenues ont donné des composantes moins corrélées que les vraies composantes  $\hat{c}^{\ell}$ .

Les matrices de corrélation entre variables communes  $R_A$  et  $R_B$  sont ici sensiblement différentes ; il est donc indispensable, pour valider la greffe, de comparer les graphiques factoriels similaires dont on dispose dans A (les  $c^l$ ) et dans B (les  $\hat{c}^l$ ).

Si on superpose les deux cercles de corrélation 1-2 (figure 5.6.a), les variables "greffées" apparaissent peu éloignées des variables initiales. En l'absence de contradictions patentes, on considère que les composantes factorielles  $\hat{c}^l$  (B) ont même interprétation que les composantes initiales  $c^l$  (A). Le graphique de la figure 5.6.a, qui valide la greffe, apporte de plus un résultat digne d'intérêt : les opinions relativement aux thèmes de conflit se structurent de la même façon dans les deux enquêtes et cela malgré certains écarts non négligeables au niveau des corrélations individuelles (cf. figure 5.5.a supra).

Sur le plan factoriel 1-2 issu de A, le point moyen respectif de l'enquête B (projeté en point supplémentaire) est assez éloigné de l'origine, où se trouve le centre de gravité des individus de l'enquête A (figure 5.6.b). Ce résultat n'a rien d'étonnant : l'échantillon de la pré-enquête est composé principalement d'enseignants et étudiants, de cadres supérieurs et de jeunes venant de la région parisienne ; dans A, les coordonnées des points moyens associés à toutes ces catégories sont bien du côté où se retrouve le point moyen de B.

L'important dans la figure 5.6.b est la disposition, presque identique, des points moyens homologues concernant les mêmes variables du signalétique (la pratique religieuse et la tendance politique). On retrouve ici la remarque de la fin du paragraphe 3.2.4. : la différenciation à l'intérieur de ce groupe particulier semble s'opérer, en gros, de la même façon que dans la population nationale.

Remarque : en principe, pour effectuer les comparaisons, il ne faut pas utiliser, dans A, les coordonnées des variables ou des individus sur les indicateurs factoriels initiaux ; on doit avoir recours aux coordonnées par rapport aux indicateurs reconstruits à partir de la formule (3.13). Du côté des variables, ceci revient à diviser les coordonnées initiales par les coefficients de corrélation multiple correspondants. Ce calcul n'a pas été effectué pour le graphique de la figure 5.6.b, d'où les corrélations plus élevées en moyenne pour les variables de B.

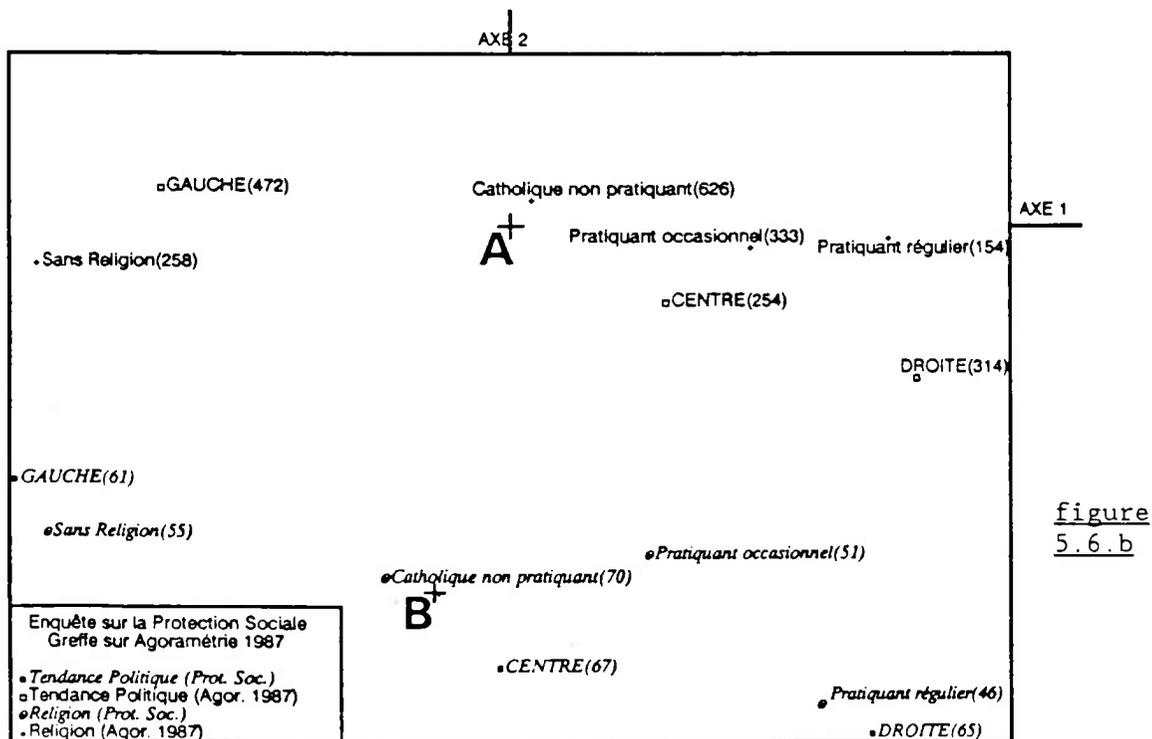
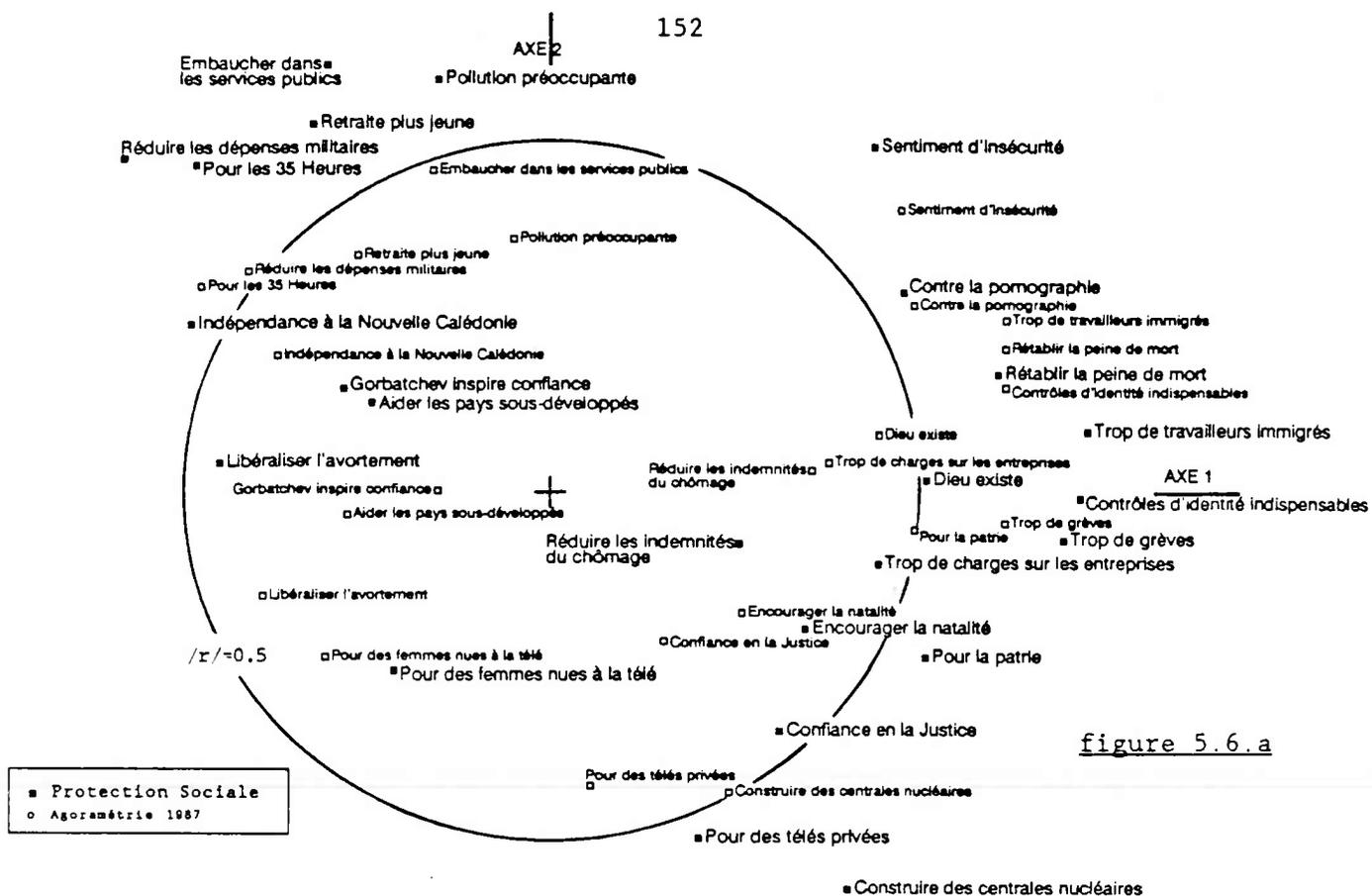
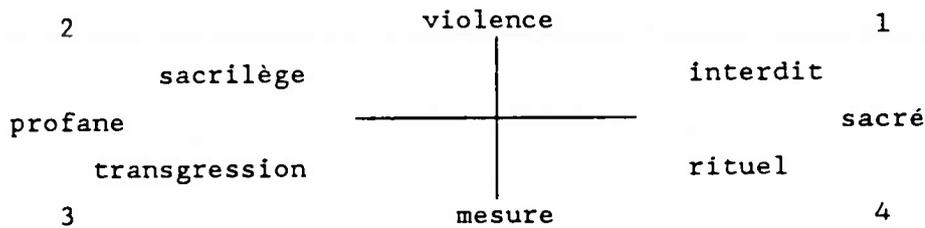


FIGURE 5.6 : Comparaison de A et B sur les cercles de corrélation et les plans principaux

## Pratique de la greffe

La technique de la greffe produit deux types de graphiques.

Un certain nombre de graphiques du premier type apparaissent à la figure 5.7. Sur le cercle de corrélation 1-2 correspondant aux composantes de référence  $\hat{c}^1$  et  $\hat{c}^2$ , on a projeté, en points supplémentaires, les variables des différentes parties du questionnaire.



Se situent du côté du sacré, comme on s'y attendait, tout ce qui évoque, de près ou de loin, la défense des valeurs traditionnelles ou de la propriété (5.7.d). Si on regarde avec un peu plus d'attention, on s'aperçoit que cela ne va pas toujours "de soi" ; remarquer les positions occupées par les variables "il faut renvoyer les immigrants" (5.7.a) et "supprimer le remboursement de l'IVG" (5.7.b). Du côté du défi (sacrilège), on retrouve bien les valeurs égalitaires et de solidarité, avec : "augmenter les cotisations patronales" et "embaucher plus d'handicapés" (5.7.b), par exemple. On est du côté de la dramatisation (violence), et là encore rien d'étonnant, avec des variables comme "il faut freiner l'automatisation" (5.7.a) ou "recul de niveau de vie" (5.7.c).

Deux graphiques du deuxième type apparaissent à la figure 5.8. Sur les cercles de corrélation issus des analyses en composantes principales effectuées sur les différentes parties du questionnaire, on envoie en points supplémentaires les composantes de référence  $\hat{c}^1, \hat{c}^2, \hat{c}^3$  considérées comme des variables illustratives.

On a contrôlé précédemment (figure 5.6) que, dans la population très particulière enquêtée, la structuration des thèmes de conflit s'effectuait de la même façon que dans la population des Français. Aussi, ces derniers graphiques présentent un intérêt majeur. En effet, après cette vérification, il est raisonnable d'admettre l'hypothèse : *la structura-*

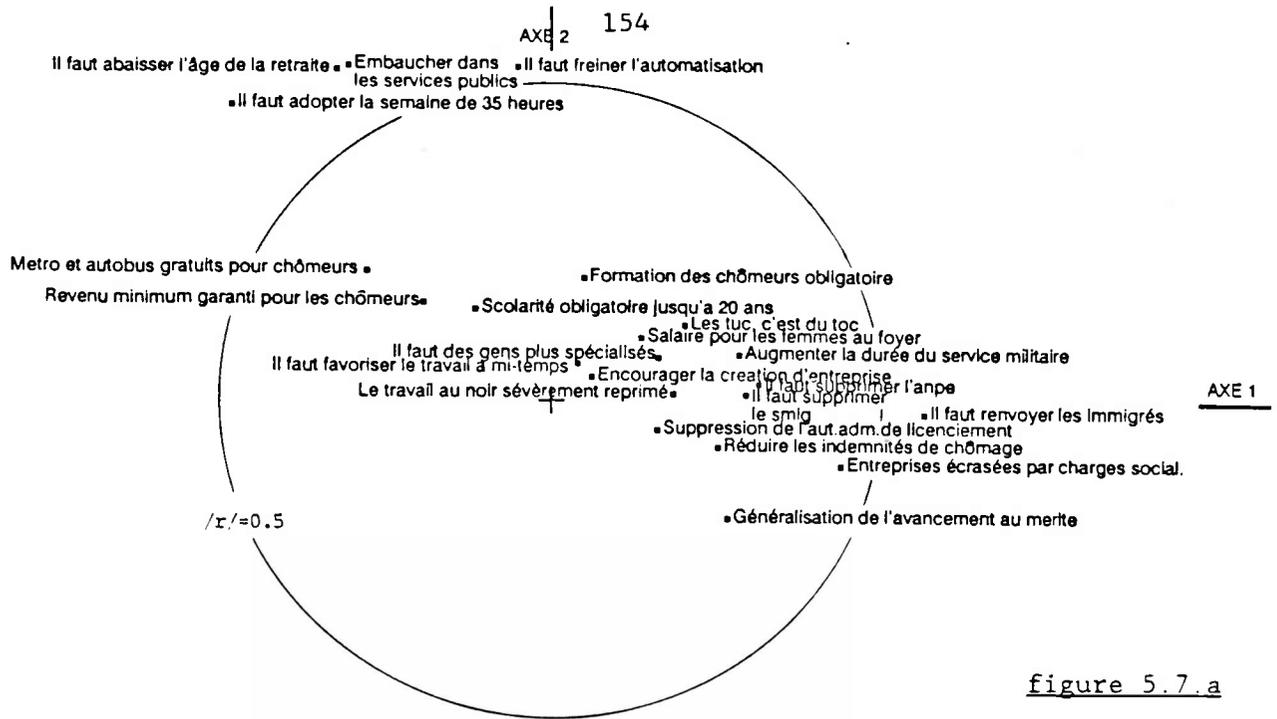


figure 5.7.a

Enquête sur la Protection Sociale  
 Greffe sur Agoramétrie 1987  
 • Pour résoudre le problème du chômage:

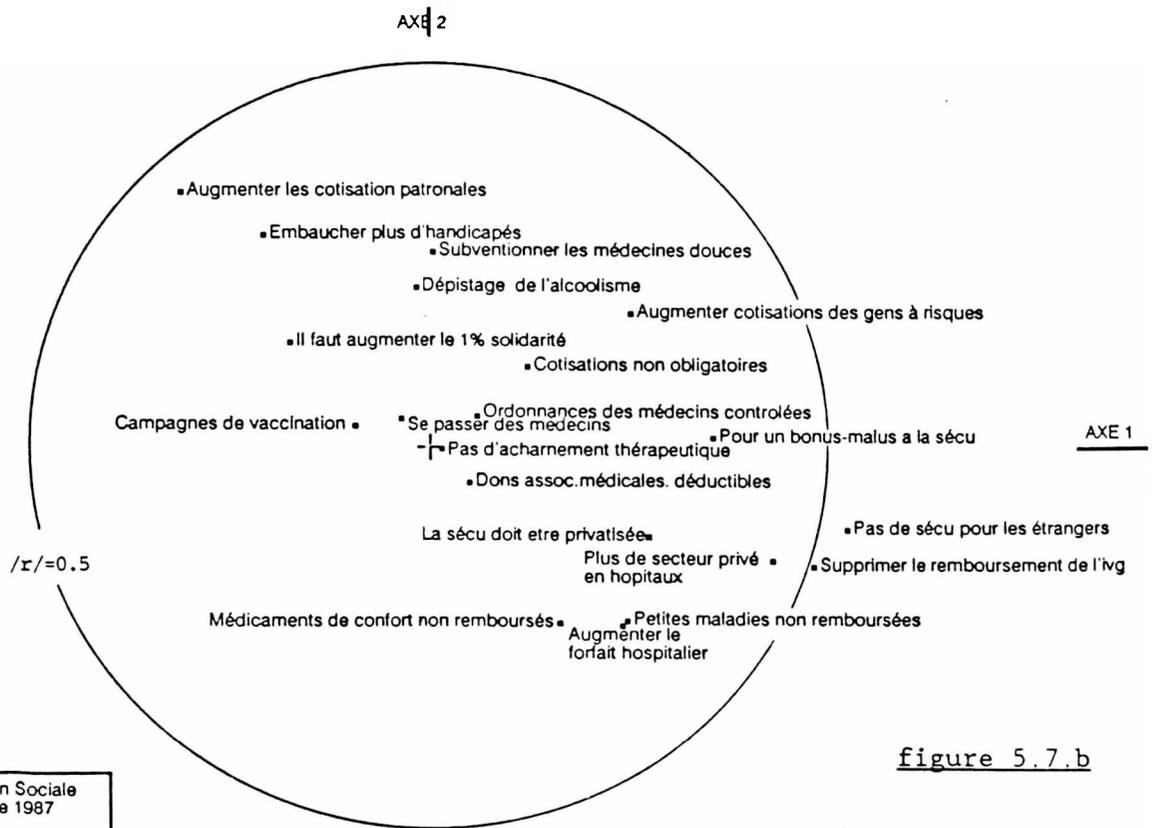


figure 5.7.b

Enquête sur la Protection Sociale  
 Greffe sur Agoramétrie 1987  
 • Thèmes sur la protection sociale

FIGURE 5.7 : Corrélations entre variables et composantes de référence (Suite...)

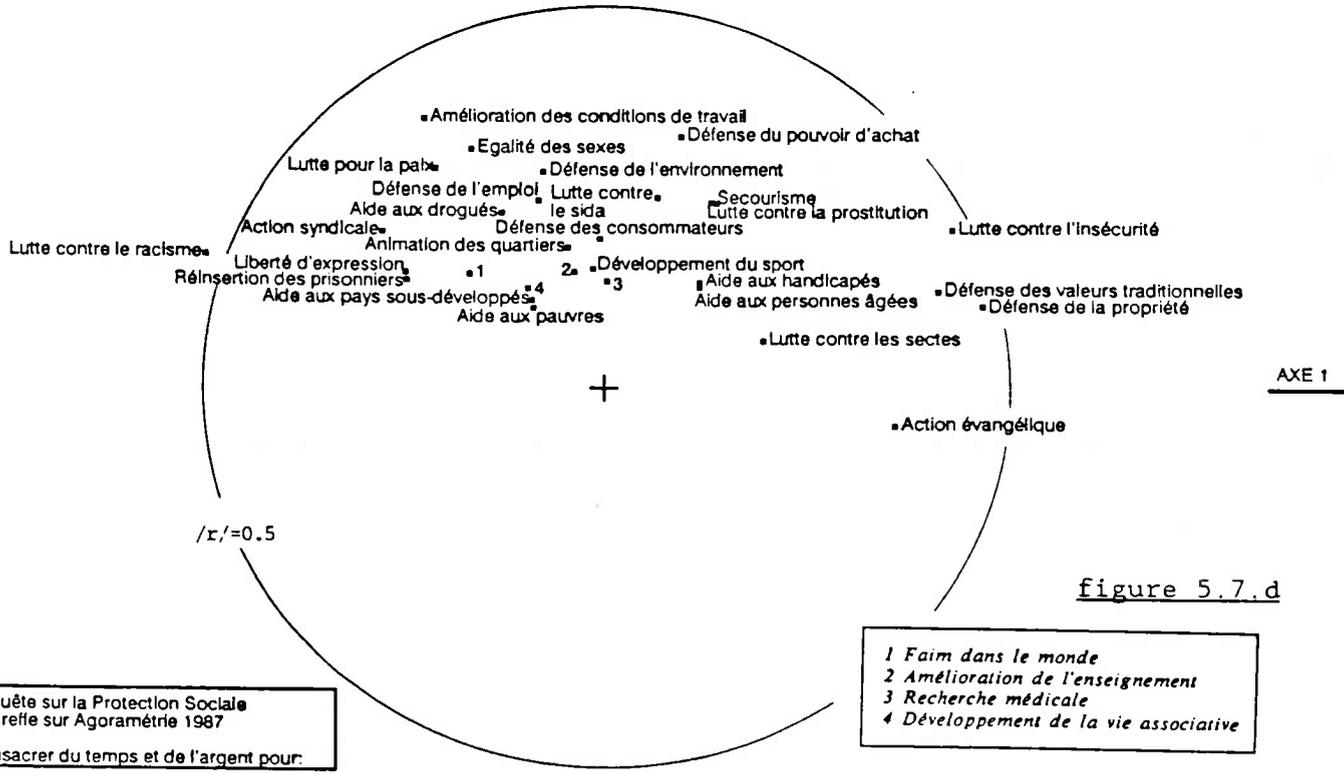
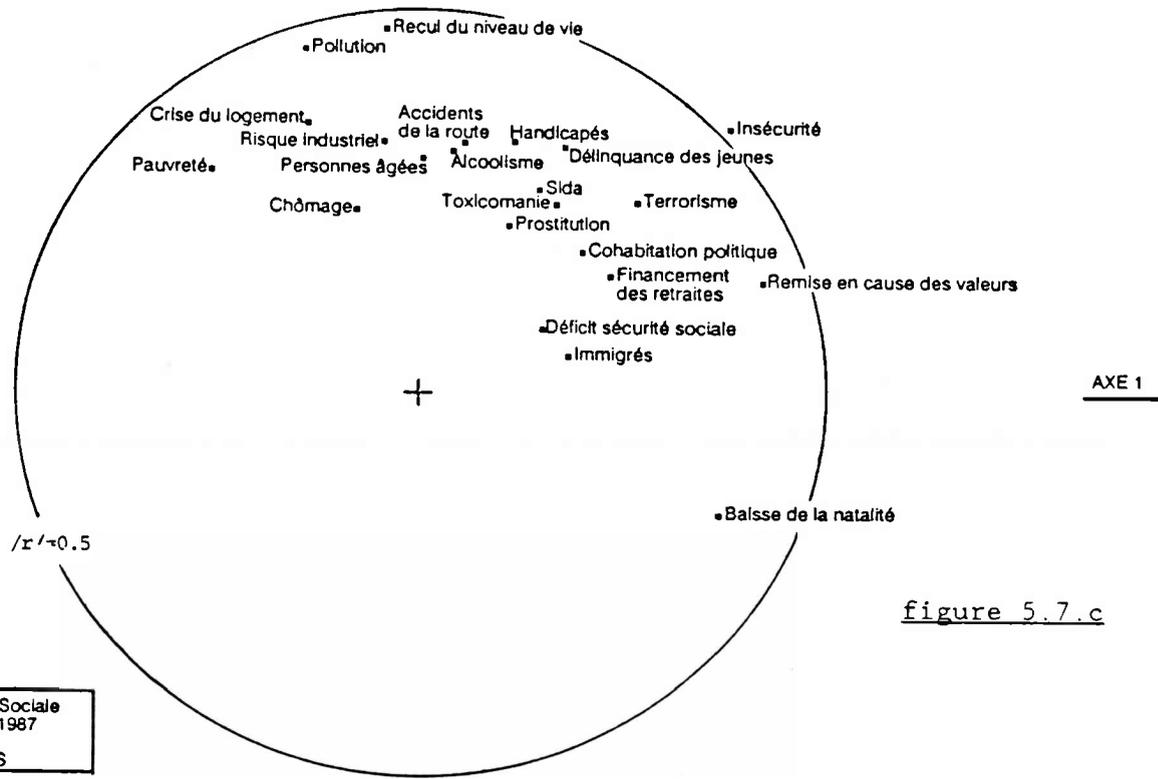


FIGURE 5.7 (... Suite)

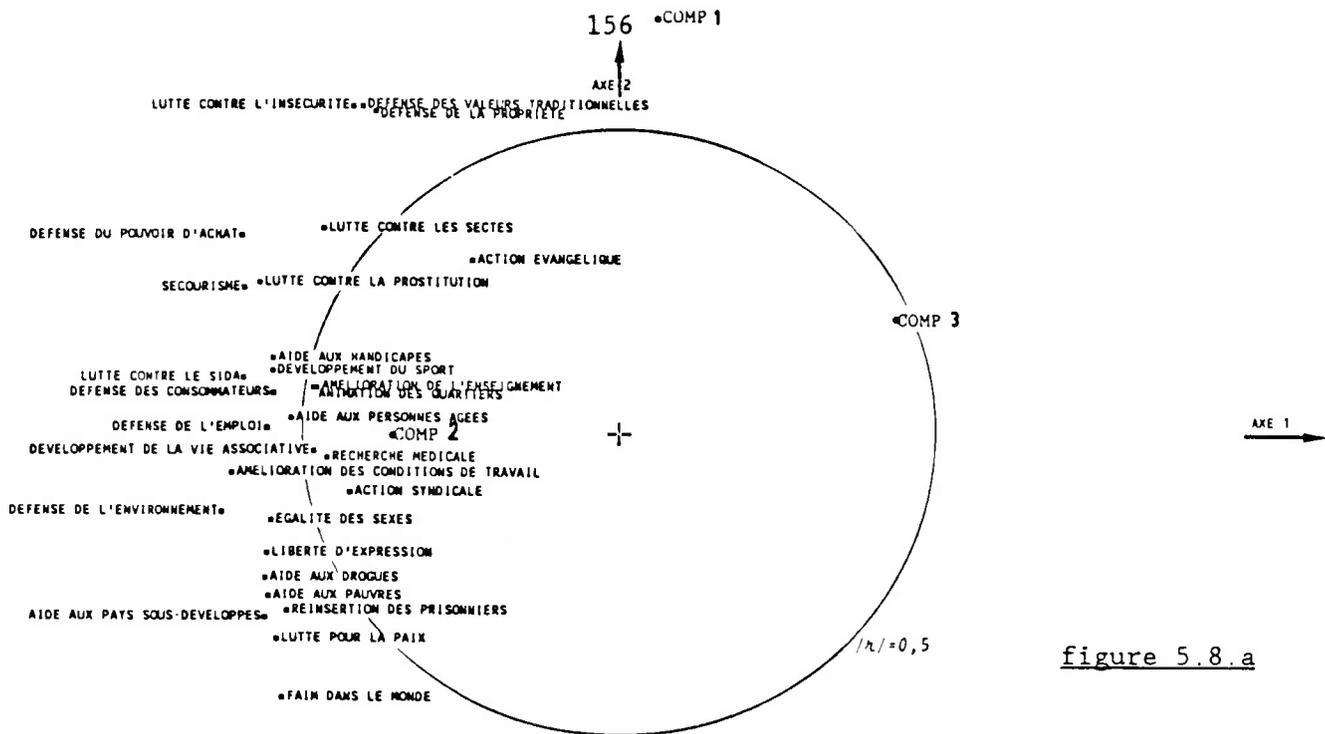


figure 5.8.a

Enquête sur la Protection Sociale  
 Consacrer du temps et de l'argent pour...  
 A.C.P.  
 Cercle des corrélations 1-2

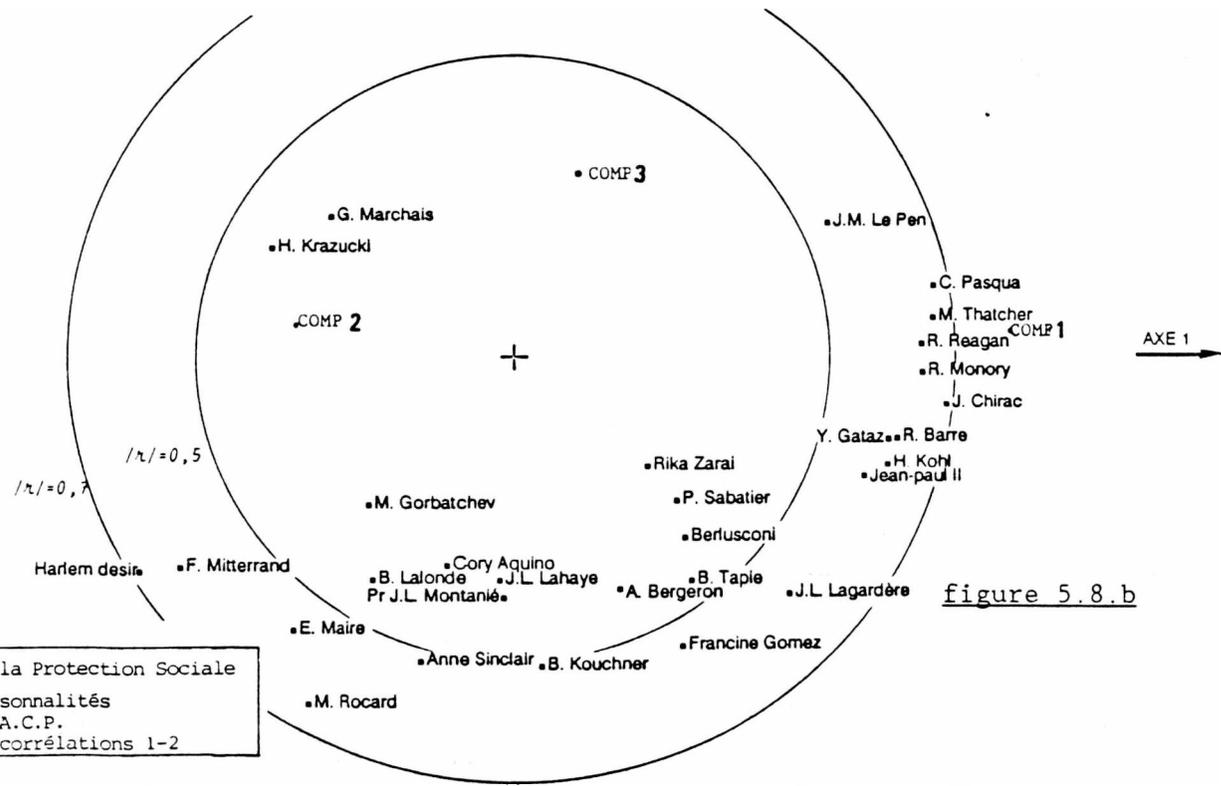


figure 5.8.b

Enquête sur la Protection Sociale  
 Personnalités  
 A.C.P.  
 Cercle des corrélations 1-2

COMP1, COMP2, COMP3 : composantes de référence d'Agoramétrie

FIGURE 5.8 : Les composantes de référence en tant que variables illustratives

tion des opinions, et cela de façon très générale, est équivalente dans les deux populations. En ce qui concerne les liaisons, les résultats fournis par la pré-enquête peuvent alors être considérés comme proches de ceux que l'on aurait obtenus en disposant d'un échantillon représentatif. En particulier, les indicateurs de référence d'Agoramétrie sont des bonnes variables illustratives pour interpréter les composantes extraites des différentes parties du questionnaire, composantes pouvant être considérées comme caractéristiques de la population générale.

Examinons le graphique de la figure 5.8.a, par exemple. On exprime son intérêt pour les actions militantes selon une première dimension qui relève à la fois de la dramatisation ( $c^2$ ) et de la béatitude ( $c^3$ ). La deuxième dimension principale, qui sépare bien les différentes actions, fait intervenir clairement des idéologies : elle est fortement corrélée à la première composante de référence (stabilité, sacré).

Le graphique de la figure 5.8.b suggère que les dimensions de référence d'Agoramétrie peuvent être produites sans passer par les thèmes de conflit. En analysant l'image d'un ensemble de personnalités variées, on retrouve, avec une bonne précision, la première dimension de référence. Il suffirait d'un ensemble de personnalités plus varié et plus vaste pour obtenir, assurément, les autres dimensions. Ce constat est cohérent avec le "schéma de l'économie des rôles" introduit (section 1.4.) ; acteurs (ici, tels qu'ils sont perçus) et individus du public sont en harmonie dans leur diversité.

#### **Bilan : ce qu'apporte la greffe sur le référent Agoramétrie**

On discute dans le grand public de la lutte contre le racisme ou contre l'insécurité, de la défense du pouvoir d'achat et pourquoi pas de l'action évangélique. On est éventuellement prêt à consacrer du temps et de l'argent à ces actions qui peuvent être considérées comme d'intérêt général. Mais quels sont les ressorts qui font que l'on s'intéresse plus à un certain type d'action ? L'argent et le temps dont on dispose ne sont manifestement pas en cause. Comme le montrent clairement les résultats (figures 5.7.c et 5.8.a) de la greffe sur le référent Agoramétrie, on touche ici aux mentalités, aux valeurs, c'est à dire aux déterminants, ancrés en nous, caractérisant les attitudes que l'on adopte de façon très générale face aux problèmes de la vie.

### 5.3. Sur l'évolution des structures de l'opinion

L'analyse en composantes principales de l'enquête effectuée par Agoramétrie en 1985 portait sur 1428 individus et 100 variables (thèmes de conflit). Quarante neuf parmi ces variables avaient été prises en compte dans l'enquête de référence faite en Mars 1981 où l'analyse en composantes principales avait porté sur 1543 individus et 76 variables. Dans les deux enquêtes, les variables sont supposées être représentatives du même univers (celui des conflits).

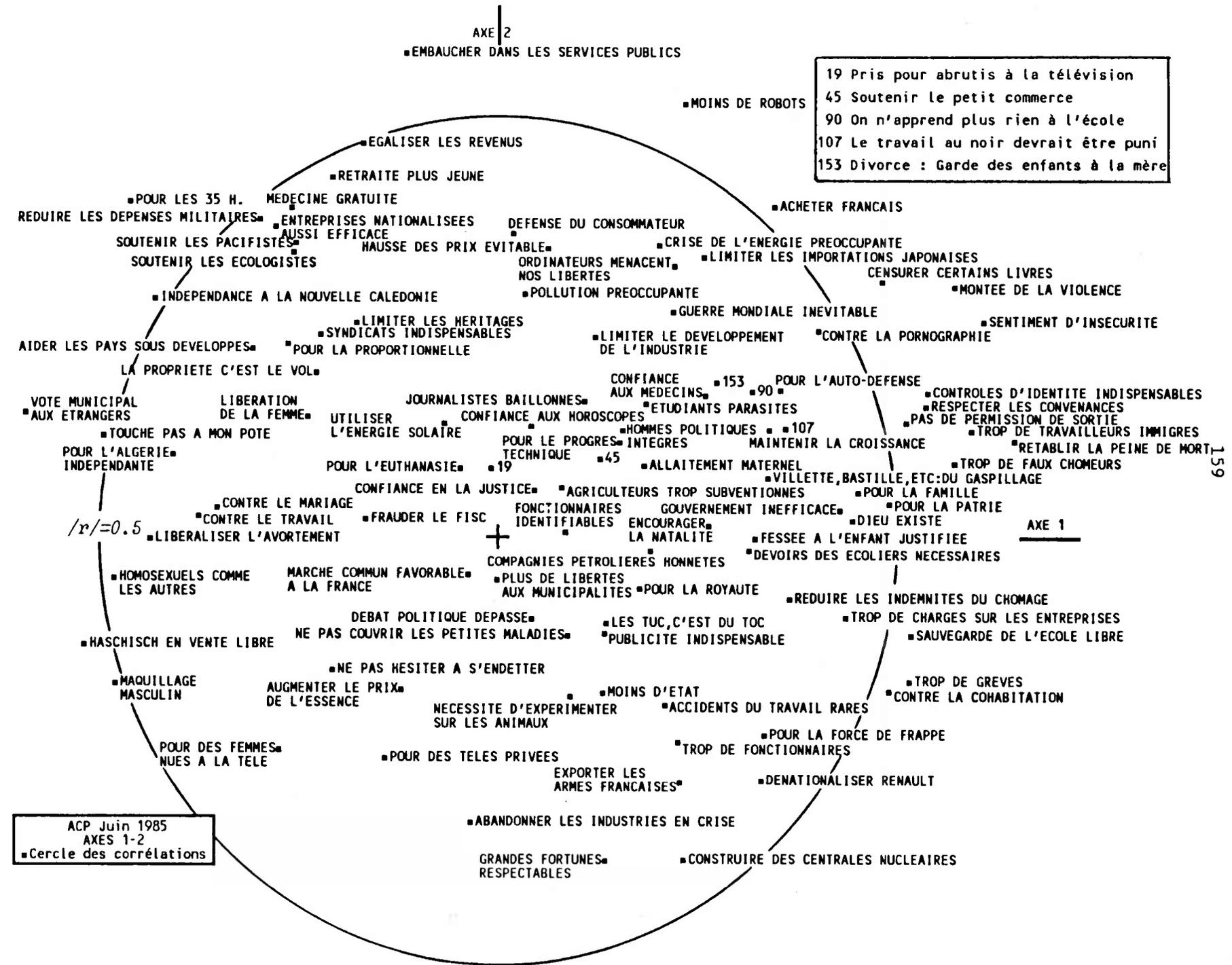
En comparant les positions des variables communes sur les cercles de corrélation (figures 5.1.a et 5.9), on constate que les composantes principales issues des deux ACP ne sont pas équivalentes. Si, en première approximation, elles se ressemblent, on constate néanmoins des différences assez importantes. Par exemple, en 1985, les thèmes du "repli" (montée de la violence, sentiment d'insécurité, trop de travailleurs immigrés) se sont déplacés : corrélation diminuée avec la deuxième composante au profit de la première. On a l'impression qu'une "rotation" des composantes s'est produit.

Mais comment a-t-on opéré pour obtenir dans l'enquête de Juin 1985 (enquête B) un ensemble de trois composantes factorielles très voisines de leurs homologues de l'enquête de Mars 1981 (enquête A) ? Sans entrer dans les détails, nous trouvons là l'occasion de décrire, dans les grandes lignes, la procédure "d'essai-erreur" proposée au paragraphe 3.2.5.

#### Construction de la passerelle

Parmi les 49 variables communes, seules les 21 les plus "proches" au sens du critère (3.6) d'égalité des matrices de corrélation ont été retenues dans la passerelle (leurs libellés réduits apparaissent dans le tableau de la figure 5.10). On sait qu'un petit nombre de variables suffit pour bâtir la matrice cible dans la procédure par rotation, et que dans les deux enquêtes, c'est à la même population d'individus (celle des Français) que l'on fait référence. Aussi, nous avons préféré éliminer d'emblée toutes les variables pour lesquelles il y avait l'ombre d'un doute. Les graphiques fournis au cours des itérations de la procédure de comparaison des matrices de corrélation (les équivalents de la figure 5.5) ne seront pas présentés ici ; mais, pour avoir une idée

FIGURE 5.9 : Le cercle des corrélations 1-2, en 1985



sur les bons et les mauvais candidats à la passerelle, on peut consulter le tableau ci-après :

A : Mars 1981 B : Juin 1985		DIEU EXISTE	LIBERAL L'AVORT	POUR LES 35 HEUR	CONFIANCE EN JUSTIC	GOUVERNEM INEFFICAC
M O Y E N N E S		A 3.18 B 3.17	3.54 3.42	3.52 3.24	2.48 2.60	3.78 3.86
CORRELATIONS		S T A B L E S			I N S T A B L E S	
DIEU EXISTE		A B	-0.32 -0.32	-0.23 -0.18	0.26 0.10	-0.14 0.12
LIBERALISER L'AVORTEMENT		A B	-0.32 -0.32	0.24 0.24	-0.11 -0.02	0.11 -0.12
POUR LES 35 HEURES		A B	-0.23 -0.18	0.24 0.24	-0.16 0.03	0.14 -0.14
CONTRE LA PORNOGRAPHIE		A B	0.30 0.28	-0.25 -0.25	-0.13 -0.15	0.18 0.09
POUR LA FAMILLE		A B	0.30 0.28	-0.15 -0.16	-0.12 -0.14	0.23 0.11
CONTRE LE MARIAGE		A B	-0.25 -0.24	0.23 0.25	0.14 0.20	-0.19 -0.13
REDUIRE LES DE- PENSES MILITAIRES		A B	-0.20 -0.17	0.17 0.16	0.24 0.25	-0.22 -0.11
CENSURER CERTAINS LIVRES		A B	0.29 0.28	-0.29 -0.30	-0.17 -0.13	0.17 0.03
RESPECTER LES CONVENANCES		A B	0.24 0.23	-0.18 -0.19	-0.13 -0.16	0.23 0.12
HOMOSEXUELS COMME LES AUTRES		A B	-0.17 -0.19	0.30 0.28	0.11 0.24	-0.09 0.00

L'examen des colonnes du tableau montre que l'on peut retenir dans la passerelle les deux premières variables dont les profils de corrélation sont extrêmement stables. Dans un premier temps, la troisième variable peut aussi être retenue. Si le changement de majorité qui a eu lieu entre les deux enquêtes n'a pas (ou a peu) affecté la signification de ces trois thèmes de conflit, il en est tout autrement pour les deux autres qui ne sauraient être retenus ; ils ne disent plus, en 1985, la même chose qu'en Mars 1981. Les corrélations avec le thème "le gouverne-

ment manque d'efficacité" se sont carrément inversées ; le thème "on peut avoir confiance en la Justice" s'est très nettement decorrélié. Remarquons que ce "glissement" dans les significations n'affecte guère les moyennes (distribution équivalente des pourcentages).

### Rotations Procrustéennes

Pour chacune des trois premières composantes principales  $a^{\ell}$  de l'enquête **A**, on cherche une combinaison linéaire  $f^{\ell} = \sum_{j=1}^5 t_{j\ell} b^j$  des composantes principales significatives normées  $b^j$  de **B** de sorte que, dans leurs enquêtes respectives, les deux composantes  $a^{\ell}$  et  $f^{\ell}$  aient les mêmes corrélations avec les 21 variables de la passerelle. Si on revient aux notations du paragraphe 3.2.5, ici :  $q=21$ ,  $k_A=3$  et  $k_B=5$ .

On notera aussi comme précédemment :

- $A(21 \times 3)$ , la matrice cible constituée des corrélations entre les variables de la passerelle et les composantes  $a^{\ell}$  dans **A**,
- $B(21 \times 5)$ , la matrice des corrélations entre variables de la passerelle et les cinq premières composantes principales  $b^j$  dans **B**,
- $F(21 \times 3)$ , la matrice des corrélations entre variables de la passerelle et composantes  $f^{\ell}$  dans **B**.

En regroupant les coefficients  $t_{j\ell}$  dans une matrice  $T(5 \times 3)$ , le problème des rotations revient à calculer la matrice "de rotation"  $T$  pour laquelle la matrice  $F=BT$  devient la plus proche possible de la matrice  $A$ . La proximité (ou l'écart) entre les deux matrices est habituellement mesurée à l'aide des indices proposés dans le paragraphe 3.2.5, l'indice le plus utilisé étant l'écart de procruste. Rappelons en la formule (écart carré moyen entre éléments homologues) :

$$e^2 = \frac{\text{trace}[(A-BT)'(A-BT)]}{k_A * q} \quad (3.30)$$

La minimisation par rapport à  $T$  de (3.30) découle de la décomposition en valeurs singulières de  $A'B$ . Cependant la matrice  $T$  cherchée ne peut pas être obtenue comme la solution d'un simple problème numérique. Dans l'équivalence à une rotation  $T$  près entre structures factorielles dans **A** et **B**, on doit partir, comme en analyse factorielle confirmatoire, d'une première hypothèse. On impose la contrainte d'orthogonalité sur  $T$

et cela en retenant toutes les variables dans la matrice cible A. On converge alors progressivement vers une solution en s'efforçant d'y voir de plus en plus clair dans la structure factorielle de B. Les critères de ressemblance entre structures et, par conséquent, le solution finale dépendent largement des considérations subjectives. Et c'est dans cet esprit qu'il faut suivre les opérations décrites ci après.

■ Précisons d'abord la dimension  $k_B$  de l'espace factoriel de B dans lequel la rotation va s'effectuer. S'agissant de retrouver  $k_A=3$  composantes, il est logique de commencer les rotations orthogonales avec  $k_B=2k_A=6$ . Pour la solution  $T_1$  obtenue, l'écart moyen entre éléments homologues de A et BT est alors égal à  $e_1=0.0506$ , valeur traduisant une qualité d'ajustement pouvant être considérée déjà comme satisfaisante.

T'T-I ; $k_B=6$ ; q=21						<u>matrice <math>T'_1</math></u>
-0.834	-0.322	-0.441	0.042	0.061	0.034	
-0.490	0.710	0.407	0.237	-0.185	0.014	
-0.242	-0.443	0.748	-0.428	0.038	0.026	

La dernière ligne de  $T_1$  ne contient pas d'éléments significativement différents de zéro. Ayant d'abord vérifié que la rotation avec  $k_B=7$  conduit à une ligne supplémentaire de zéros dans T,  $k_B$  est décrémenté :  $k_B=5$ . La qualité d'ajustement alors obtenue n'est guère détériorée ( $e_2=0.0507$ ), les coefficients de rotation significatifs restant pratiquement inchangés. Cette fois, toutes les lignes de la nouvelle matrice de rotation  $T_2$ , contiennent au moins un élément non nul. Aussi,  $k_B$  est-il pris égal à cinq.

T'T-I ; $k_B=5$ ; q=21					<u>matrice <math>T'_2</math></u>
-0.835	-0.327	-0.438	0.045	0.049	
-0.490	0.708	0.407	0.238	-0.192	
-0.240	-0.446	0.749	-0.426	0.029	

■ Examinons maintenant la décomposition de l'écart de procruste  $e_2^2$  par axe et par variable comme le propose le programme informatique (fi-

gure 5.10.b) : sur les bords du tableau des différences entre éléments homologues des matrices A et BT, on fait apparaître la contribution (en pour cent) de chaque variable et de chaque composante à l'écart procruste. Pour les deux premières composantes, ces différences sont illustrées graphiquement dans la figure 5.10.a, où les deux cercles de corrélation  $(f^1, f^2)$  et  $(a^1, a^2)$  ont été superposés. Cette "analyse de la variance" permet, par exemple, de voir immédiatement que l'ajustement s'effectue mieux pour la deuxième composante et que la moitié de l'écart total est due à trois ou quatre variables seulement. Diverses vérifications, effectuées, théoriquement, en faisant des aller et retour avec l'analyse précédente, viennent compléter l'étude de la stabilité des résultats ; les principales d'entre elles sont exposées ci-après.

■ Si une composante de référence est vraiment caractéristique, elle doit pouvoir être retrouvée indépendamment des autres composantes. C'est pourquoi les vérifications commencent habituellement en s'assurant que la contrainte d'orthogonalité sur T n'est pas trop restrictive. La relaxation de cette contrainte conduit à des "rotations" obliques : chaque colonne  $t_j$  de T est calculée indépendamment, sous la seule contrainte :  $t_j' t_j = 1$ .

Les résultats de toutes les rotations obliques effectuées ont confirmé l'à propos de l'hypothèse d'orthogonalité entre les trois composantes  $f^l$ . Considérons, par exemple, la matrice  $T_3$  obtenue avec toutes les  $q=21$  variables dans la matrice cible et avec  $k_B=6$  composantes  $b^j$  (quand on fait des rotations obliques, la sixième composante principale  $b^6$  semble intervenir légèrement) :

diag(T'T)=I ; $k_B=6$ ; $q=21$						
-0.800	-0.290	-0.509	0.063	0.072	0.086	<u>matrice <math>T_3'</math></u>
-0.456	0.762	0.310	0.271	-0.178	0.104	
-0.220	-0.454	0.738	-0.442	0.064	0.027	

La matrice  $T_3$  est voisine de la matrice orthogonale  $T_1$ . Les composantes obliques sont peu corrélées entre elles :

$$T_3' T_3 = \begin{matrix} & \begin{matrix} 1.000 & 0.000 & -0.089 \\ 0.000 & 1.000 & -0.145 \\ -0.089 & -0.145 & 1.000 \end{matrix} \end{matrix}$$

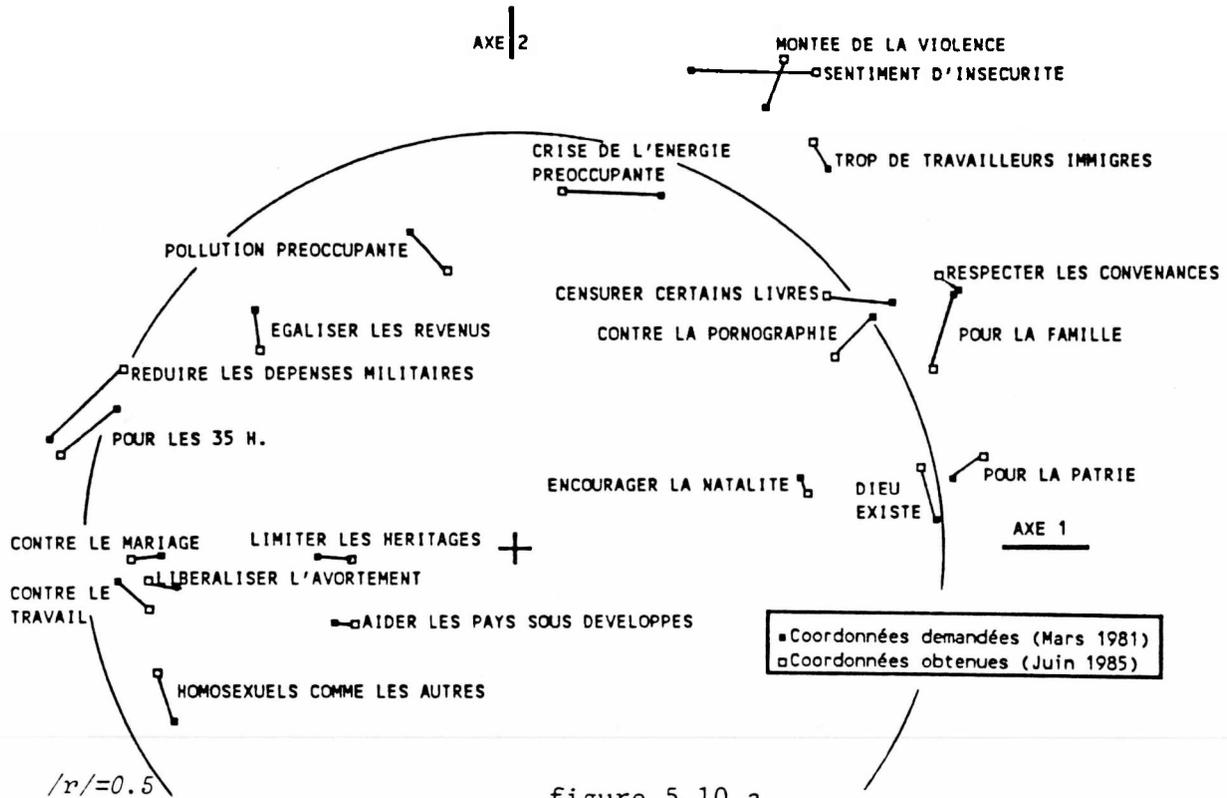


figure 5.10.a

VARIABLES DE LA PASSERELLE	AXE 1	AXE 2	AXE 3	%
LIBERALISER L'AVORTEMENT	0.032	-0.007	-0.023	0.97
DIEU EXISTE	0.018	-0.066	-0.084	7.24
EGALISER LES REVENUS	-0.005	0.047	-0.041	2.42
TROP DE TRAVAILLEURS IMMIGRES	0.018	-0.033	0.005	0.90
LIMITER LES HERITAGES	-0.039	0.003	0.005	0.98
CONTRE LA PORNOGRAPHIE	0.045	0.047	0.009	2.65
AIDER LES PAYS SOUS DEVELOPPES	-0.024	0.002	-0.010	0.42
POLLUTION PREOCCUPANTE	-0.044	0.046	-0.064	5.03
POUR LA PATRIE	-0.036	-0.027	-0.011	1.31
ENCOURAGER LA NATALITE	-0.008	0.019	0.038	1.16
POUR LA FAMILLE	0.020	0.083	-0.007	4.54
CONTRE LE TRAVAIL	-0.037	0.034	-0.143	14.18
CONTRE LE MARIAGE	0.035	0.004	0.044	1.97
REDUIRE LES DEPENSES MILITAIRES	-0.085	-0.085	0.078	12.63
CENSURER CERTAINS LIVRES	0.078	-0.008	0.024	4.16
RESPECTER LES CONVENANCES	0.023	-0.018	0.032	1.17
SENTIMENT D'INSECURITE	-0.146	0.002	0.005	13.25
CRISE DE L'ENERGIE PREOCCUPANTE	0.115	-0.006	-0.020	8.44
POUR LES 35 H.	0.064	0.056	0.038	5.31
MONTEE DE LA VIOLENCE	-0.021	-0.060	-0.094	7.88
HOMOSEXUELS COMME LES AUTRES	0.019	-0.059	0.040	3.38
contribution des composantes (%) :	40.41	24.14	35.45	100.0%

figure 5.10.b

FIGURE 5.10 : Décomposition de l'écart procruste  $e_2^2$

et les corrélations entre composantes obliques et orthogonales du même ordre (éléments diagonaux de  $T_1'T_3$ ) sont égales à : 0.995, 0.989, 0.998.

■ Etant donné que le nombre de variables actives dans l'ACP de B est supérieur à celui dans A (100 contre 76), il est raisonnable de s'interroger sur la présence d'un effet de "dilatation" : l'ajustement est-il amélioré si on accepte que les corrélations dans la matrice F sont en moyenne inférieures (proportionnelles à un coefficient de dilatation) à celles dans la matrice A ? Deux essais de rotation orthogonale avec dilatation ont été effectués : dans le premier, où on a introduit un seul coefficient de dilatation global d, on a trouvé  $d=0.971$  ; dans le second, avec un coefficient de dilatation séparé pour chaque composante, on a trouvé :  $d_1=0.968$ ,  $d_2=0.990$ ,  $d_3=0.917$ . Dans les deux cas, la matrice de rotation obtenue était pratiquement identique à la matrice  $T_1$  (corrélations entre composantes du même ordre supérieures à 0.999). On retient donc la solution sans dilatations.

La qualité d'ajustement de la solution orthogonale initiale (matrice  $T_2$ ) étant déjà satisfaisante, nous n'aurons pas besoin de revenir ultérieurement quand on changera de cible, ni sur les rotations obliques, ni sur les coefficients de dilatation.

■ Cherchons maintenant à affiner la solution orthogonale sans dilatations avec  $k_g=5$  composantes. A la fin de chaque essai, le programme informatique fournit l'équivalent du tableau des écarts de la figure 5.10.b ; il affiche, en même temps, à l'écran les cinq plus grands écarts triés par ordre d'importance. Pour la solution correspondant à la figure 5.10.b ces cinq écarts sont :

	Axe	valeur demandée( $a_{ij}$ )	valeur obtenue( $f_{ij}$ )
Sentiment d'insécurité	1	0.204	0.350
Contre le travail	3	-0.018	0.125
Crise de l'énergie préoccupante	1	0.167	0.052
Montée de la violence	3	-0.033	0.061
Réduire les dépenses militaires	2	0.131	0.216

Plusieurs rotations orthogonales sont alors effectuées. A chaque fois, on réduit l'influence dans la rotation des éléments qui correspon-

dent aux plus grands écarts en leur affectant une valeur intermédiaire entre la valeur demandée  $a_{ij}$  et la valeur obtenue après rotation  $f_{ij}$  ; parfois en les ôtant directement de la matrice cible, quand on n'a pas de raisons particulières pour tenir à la valeur initiale  $a_{ij}$  demandée (par exemple, on retire la coordonnée de la variable "contre le travail" sur le troisième axe). Dans tous les cas, la stabilité de la matrice T qui a été obtenue est remarquable. Donnons, à titre d'exemple, la matrice  $T_4$  correspondant à la rotation orthogonale où les quatre variables qui contribuent le plus à l'écart de procruste  $e_2^2$  ont été écartées simultanément de la matrice cible :

T'T=I ; $k_B=5$ ; $q=17$					
-0.862	-0.277	-0.417	0.077	0.029	matrice $T'_4$
-0.437	0.758	0.426	0.198	-0.120	
-0.252	-0.454	0.748	-0.412	0.026	

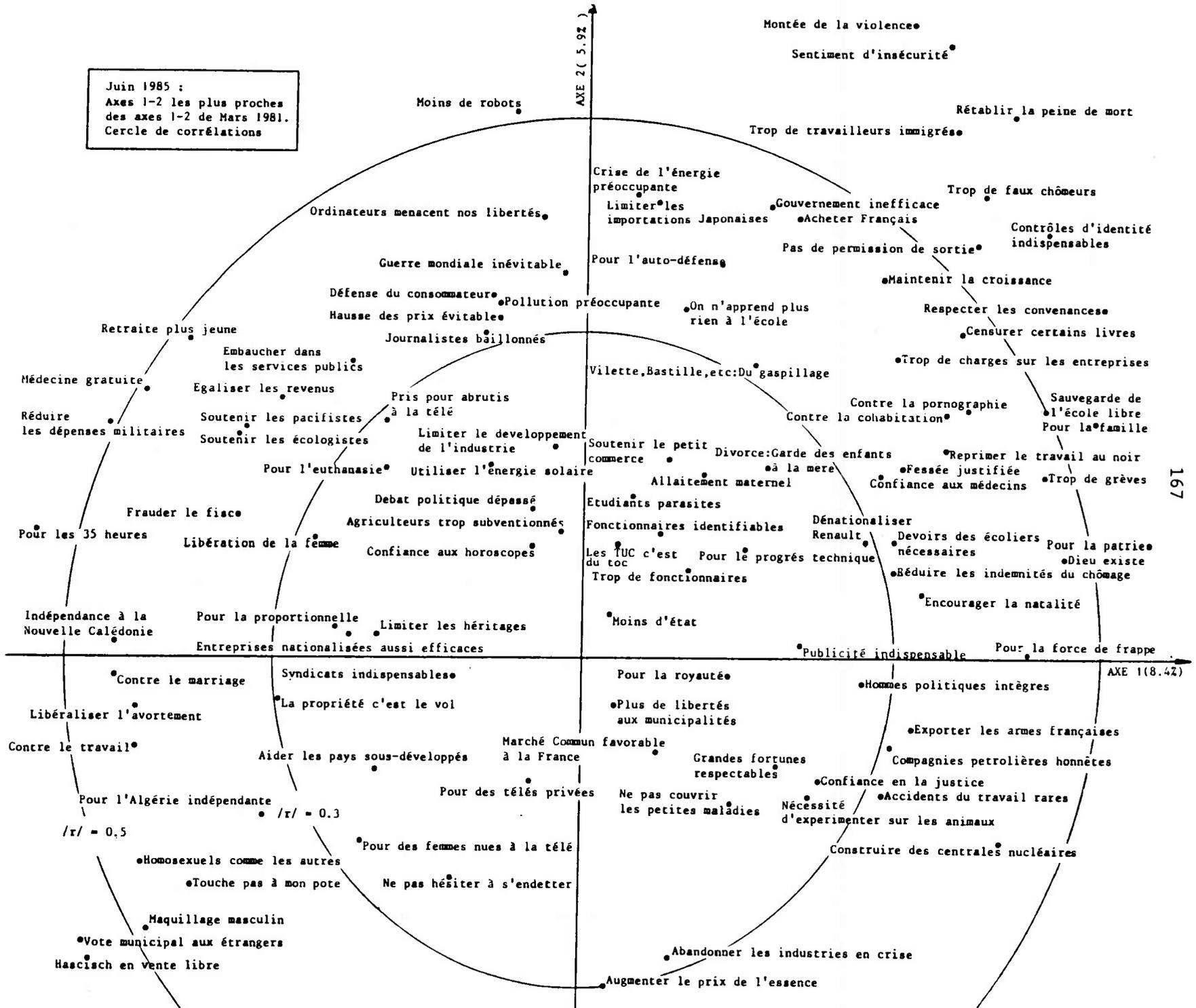
Certes, en l'absence des quatre variables mentionnées, l'écart moyen  $e_4=0.0411$  est nettement inférieur à l'écart  $e_2=0.0507$ . Toutefois la matrice  $T_4$  reste proche de la matrice  $T_2$  ; les corrélations entre composantes du même ordre (les termes diagonaux de  $T'_4T_2$ ) valent : 0.997, 0.994 et 0.999.

Parmi les solutions  $T_2$  et  $T_4$  laquelle retenir ? La réponse est entièrement subjective. Avec la matrice  $T_4$  l'ajustement est meilleur, l'effet de rotation étant aussi moins ample pour les deux premières composantes. Avec  $T_2$  les résultats sont plus stables, un plus grand nombre de variables intervenant dans la cible. Ayant l'impression que les variables qui contribuent le plus à l'écart de procruste  $e_2^2$  ne font pas "dévier" les composantes après rotation d'une position, qui, il nous semble, est proche des composantes de Mars 1981, nous avons préféré la matrice  $T_2$ . Le cercle des corrélations 1-2 correspondant (composantes  $f^1$  et  $f^2$ ) est présenté dans le figure 5.11.

#### Deux remarques

- Si les 49 variables communes avaient été initialement retenues dans la matrice cible, on n'aurait pu exploiter correctement l'information apportée par le tableau des écarts et l'analyse de la variance.

FIGURE 5.11 : Cercle de corrélation après rotation



Prises avec les autres, les variables : "gouvernement inefficace" et "confiance en la Justice", conduisent à une rotation bâtarde où toutes les variables sont mal ajustées ; aussi bien celles qui ont effectivement changé de signification que celles qui sont restées stables. Etant alors difficile de les distinguer, on n'écarte pas forcément, à chaque étape, les bonnes variables pour converger vers la solution finale.

- La stabilité des résultats par rapport aux variables qui interviennent dans la matrice cible, prouve que l'effet global qui caractérise l'évolution des structures factorielles entre les deux enquêtes peut être assimilé à une rotation orthogonale. Nous ne saurions en dire plus sur cette évolution des structures.

Avec une procédure comme celle de Jöreskog (cf 3.1.2.) on risquerait de se perdre dans la masse des calculs en évaluant l'impact sur la fonction de vraisemblance des nombreuses modifications mineures apportées au niveau des paramètres individuels. Compte tenu de l'imprécision qui caractérise les variables d'opinion, il est clair qu'on n'est pas ici à 0.01 ni même à 0.05 près au niveau des coefficients de corrélation. Ajoutons que la minimisation numérique de la fonction de vraisemblance, sous la contrainte d'une décomposition de type ACP de la matrice de variance qui est (100×100), paraît impraticable. En effet, la majorité des paramètres dans cette décomposition doivent être estimés (paramètres "libres" dans la terminologie de Jöreskog), les seuls paramètres "fixés" étant les coordonnées des 21 (ou mieux 49, car chez Jöreskog il n'y a pas d'élimination préalable) variables de la passerelle par rapport aux trois premières composantes factorielles.

#### Commentaires sur les résultats

- La solution finale paraît d'autant plus satisfaisante que des variables communes qui ne sont pas intervenues dans la rotation, et qui peuvent être considérées comme relativement stables se retrouvent après rotation, en 1985, au même endroit du graphique qu'en 1981 (comparer figures 5.1.a et 5.11). Il en est ainsi pour la variable "retraite plus jeune" dans le quadrant du défi et la variable "maintenir la croissance" dans le quadrant du repli. Il en est de même aussi pour la variable "rétablir la peine de mort" qui se trouve en 1985 dans le quadrant du

défi, à l'endroit opposé de la variable "supprimer la peine de mort" qui se trouvait, en 1981, dans le quadrant de l'aventure.

■ Les thèmes du repli (sentiment d'insécurité, montée de la violence...) qui semblaient légèrement instables occupent, après rotation, la même position sur le cercle des corrélations qu'en 1981. Aucun changement de signification ne s'est produit comme le laissait entendre la figure 5.9.a ; une légère "sacralisation", qui s'est confirmée dans les années suivantes, est cependant mesurable. Soulignons que les coordonnées des deux thèmes sont les mêmes, qu'ils soient ou non dans la matrice cible.

■ En comparant les plans principaux des deux enquêtes on a l'impression que les groupes sociaux ont peu évolué dans leurs discours. Considérons par exemple les différentes catégories de la variable "proximité politique" : les positions relatives des partis politiques restent, en 1985 (figure 5.12.a), en gros, les mêmes qu'en Mars 1981 (figure 5.1.a).

Or cette impression est largement infirmée par les tris croisés entre la variable "proximité politique" et les thèmes de conflit. Dans le cas des partis de gauche, par exemple, on sait que depuis 1981 le discours des sympathisants du parti socialiste s'est considérablement démarqué de celui des sympathisants du parti communiste en ce qui concerne la majorité des thèmes "égalitaristes" situés dans le quadrant du "défi" : en 1985, 40% des pro-socialistes restent entièrement d'accord avec la proposition "il faut diminuer au maximum l'écart entre les revenus", contre 61% des pro-communistes ; 36% seulement parmi les premiers considèrent encore qu'"il faut abaisser l'âge de la retraite", contre 51% parmi les derniers.

En dressant le plan factoriel 1-2 (figure 5.12.b), dual du cercle des corrélations ( $f^1, f^2$ ) de la figure 5.11, le mouvement devient manifeste : on constate qu'effectivement la parti socialiste, encore au pouvoir à l'époque, a abandonné le quadrant du défi (c'est-à-dire le discours égalitariste), y laissant seul le parti communiste. Sur ce même plan, les positions des partis de la droite, et surtout de l'extrême droite, reflètent bien l'évolution qui s'est effectuée vers la dramatisation : on insiste plus, en 1985, sur les thèmes de la violence et de l'insécurité.

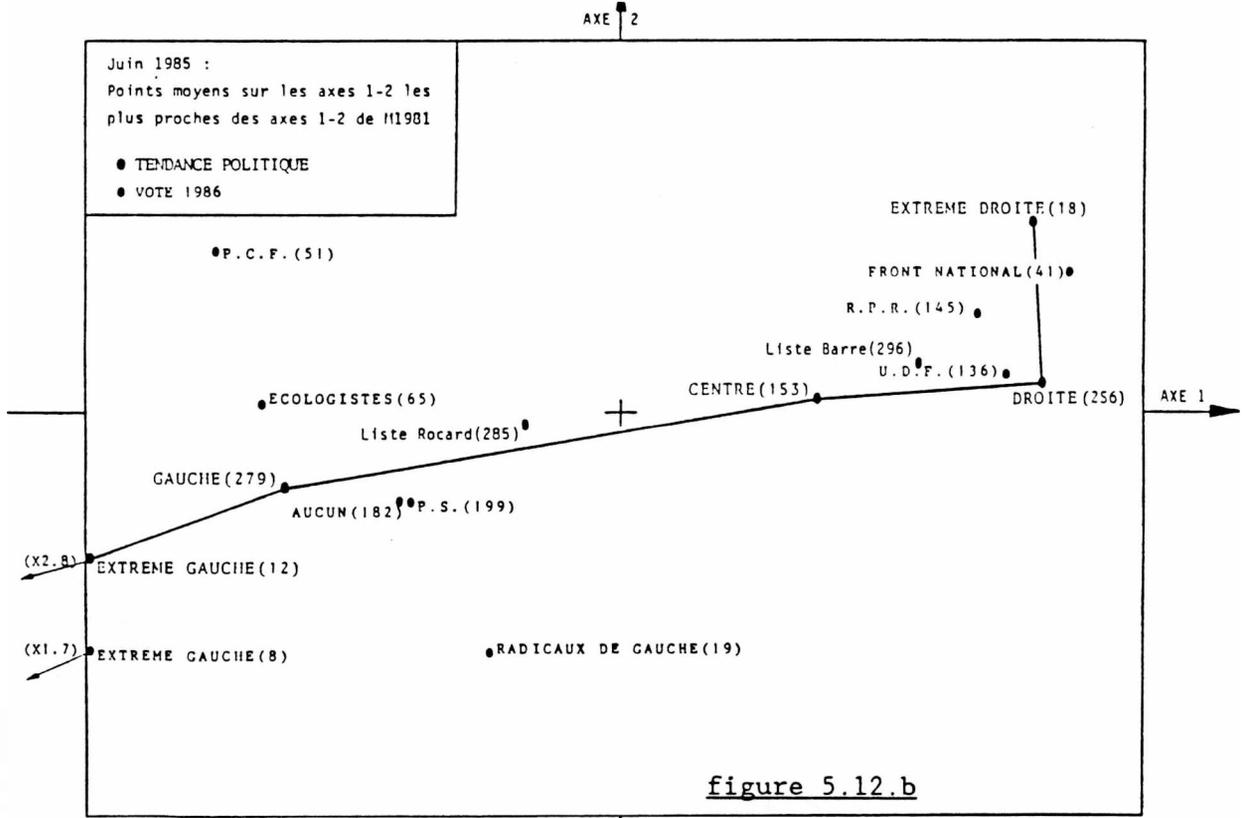
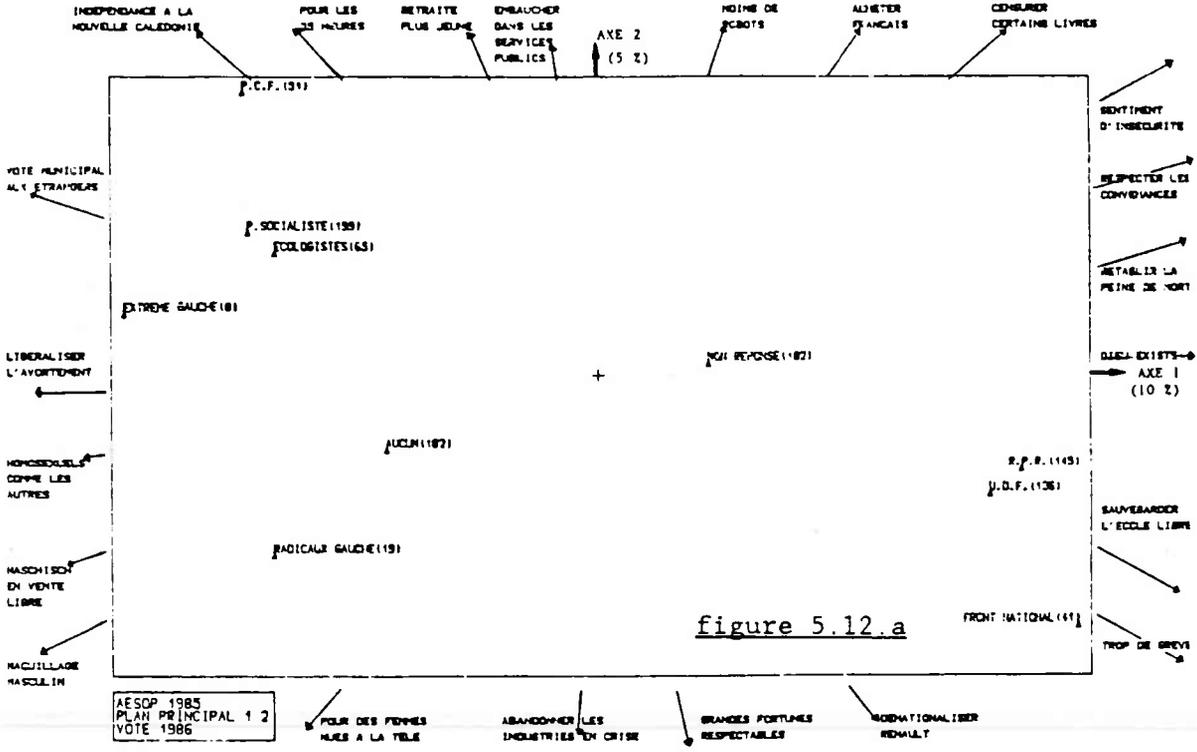


FIGURE 5.12 : Comparaison de plans factoriels avant (a) et après (b) rotation

■ Cette possibilité de retrouver, après rotation en 1985, les mêmes structures factorielles qu'en 1981, ne signifie en rien que les structures de l'opinion publique sont condamnées à l'immobilité. Depuis 1986, le recalage des dimensions factorielles sur l'enquête de Mars 1981 s'effectue moins bien. Tout ce que le cercle des corrélations 5.11 suggère est que des faits majeurs comme l'arrivée au pouvoir de la gauche en 1981, s'ils provoquent le changement (pourcentages), ne peuvent guère dans une brève période provoquer une réorganisation profonde des structures.

**Bilan : ce que montre l'analyse des évolutions avec la méthode par rotations**

Les grands faits marquants qui secouent nos sociétés ne changent pas radicalement, dans une brève période, la place que l'on occupe ou que l'on veut occuper par rapport à autrui. Cette stabilité des positions relatives se retrouve, de façon très générale, au niveau des groupes sociaux. Ce qui change ce sont les manières de dire ; celles-ci évoluent au gré de ces débats où les antagonismes précédents s'expriment. Le thème porteur de la différenciation d'hier tombe dans l'oubli ; il est remplacé par d'autres thèmes qui vont à leur tour cristalliser les passions. Ce mouvement perpétuel<sup>1</sup>, caractéristique de nos démocraties, qui s'alimente des faits d'actualité (environnement dans le schéma de l'économie des rôles de la section 1.4.) et qui les crée en retour, peut se faire sans à-coups ; il y a alors parfaite stabilité des "structures". Tout fait exceptionnel conduisant à remettre en cause le "statut" des principaux acteurs sociaux, et qui donc touche à l'organisation sociale dans sa partie la plus immédiatement communicable, est producteur de désordre ; il y a, et cela jusqu'à un certain point, rupture dans les manières de dire et donc perturbation au niveau des structures. Cette période de faiblesse favorise-t-elle l'émergence des extrémismes ? L'extrémisme de droite profiterait de la circulation de la gauche du défi à l'aventure... Quoi qu'il en soit, la machine sociale ne retrouvera sa sérénité que lorsque les incohérences apparues dans ce qui est dit disparaîtront ; il faut pour cela que le dialogue se rétablisse autour de tout un cortège d'idées qui jouisse d'une double qualité : l'harmonie synchronique (consensus entre groupes sociaux sur les grandes lignes et les formes du dialogue) ; l'harmonie diachronique (non décalage entre les idées sur lesquelles on débat et les réalités du temps).

---

<sup>1</sup>Charbonnier G. (1981). *Entretiens avec Levi-Strauss*. Plon, coll. 10/18.

#### 5.4. Structures de l'opinion et perception des risques

L'enquête sur la perception des risques<sup>1</sup> a été réalisée en Février 1986 auprès d'un échantillon à peu près représentatif de la population française (habitat rural sous-représenté). A travers les thèmes de conflit présents dans le questionnaire, il est donc possible, ici, de se greffer sur le repère factoriel de l'enquête Agoramétrie de Juin 1985 sans trop de difficultés. Ce ne sont pas les deux premières dimensions principales extraites en 1985, mais les deux premières dimensions factorielles issues de l'enquête de Mars 1981, telles qu'elles sont reconstruites dans l'enquête de 1985, qui vont servir de repère de référence. Rappelons qu'elles ont été obtenues (exemple précédent, figure 5.11) par rotations procrustéennes.

#### Questionnaire

Le questionnaire comprend environ trois cents questions réparties en cinq grandes parties. D'abord, un signalétique élargi : variables socio-démographiques et socio-culturelles, mais aussi questions factuelles ou même d'opinion portant sur des sujets très divers concernant la santé, la perception de la sécurité ou le comportement sécuritaire, le cadre de vie... Puis, quatre parties homogènes (réponses sur des échelles à cinq paliers), où l'interviewé donne successivement son avis sur la fréquence d'une trentaine de causes de mortalité (cancer du poumon, accidents de la route,...), sur les dangers présentés par cinquante deux situations à risque (voiture, drogue,...), sur sa confiance en l'organisation de la sécurité dans dix-neuf domaines (transport des matières dangereuses, banques,...) et, enfin, puisque greffe oblige, sur vingt cinq propositions référendum (les thèmes de conflit) issus du questionnaire de l'enquête Agoramétrie de Juin 1985.

---

<sup>1</sup> Pour une présentation globale des résultats de l'enquête, on pourra consulter :

Bastide S. & J.P. Pagès (1987). *Perception des risques et communication. Approche et premiers résultats*. IN : La société vulnérable. Evaluer et maîtriser les risques. Presse de l'École Normale Supérieure, 93-110.

## Objectif de la greffe

Une façon d'aborder le risque que l'on perçoit à propos d'une activité est de considérer que celui-ci découle :

- d'une information que l'on a reçue (nombre de morts par accident de la route, nombre de suicides...),
- d'une expérience que l'on a acquise ayant été en relation directe avec le risque (apprentissage),
- de certaines caractéristiques de la situation à risque abordée (pour soi ou pour les autres, choisi ou subi, naturel ou artificiel...) ou de certaines qualités qui nous sont propres (on est anxieux de nature).

Pour le groupe de travail "Perception des Risques et Communication", toutes ces explications "techniques" étaient par trop insuffisantes. Conduisant à établir des relations de cause à effet qui paraissent évidentes<sup>1</sup>, elles occultaient le fait que le risque est au centre d'un conflit social : la perception sociale ne renvoie pas à des caractéristiques objectives explicatives, mais aux représentations que l'on a du risque, représentations qui se construisent avec le débat social qu'il provoque<sup>2</sup>. On a donc glissé, ici, de la raison et du technique à l'imaginaire, au symbole et à la métaphore. Au delà de ce qu'il ressent au niveau de son corps, que dit alors l'individu quand il s'exprime sur le danger ? L'affirmation qu'il existe un danger, n'est-elle pas un moyen pour justifier une prise de position qui trouve son origine ailleurs ? L'expression sur le danger n'est-elle pas, finalement, qu'un discours ?

C'est bien pour répondre à ces questions que l'on a cherché à se replacer sur le repère de référence des indicateurs globaux d'opinion d'Agoramétrie. Afin de réaliser au mieux la greffe, on a inséré dans le questionnaire les dix thèmes "optimaux" dans la reconstruction des deux premières composantes ; quinze autres thèmes considérés comme ayant un rapport avec le risque sont venus les rejoindre. Ici, nous ne présentons qu'une partie des résultats, celle concernant les activités dangereuses.

---

<sup>1</sup>nous sommes ici en plein dans l'approche "mécaniste" du phénomène de l'opinion décrite à la section 1.2

<sup>2</sup>on retrouve ici l'approche "constructionniste" décrite à la section 1.3.

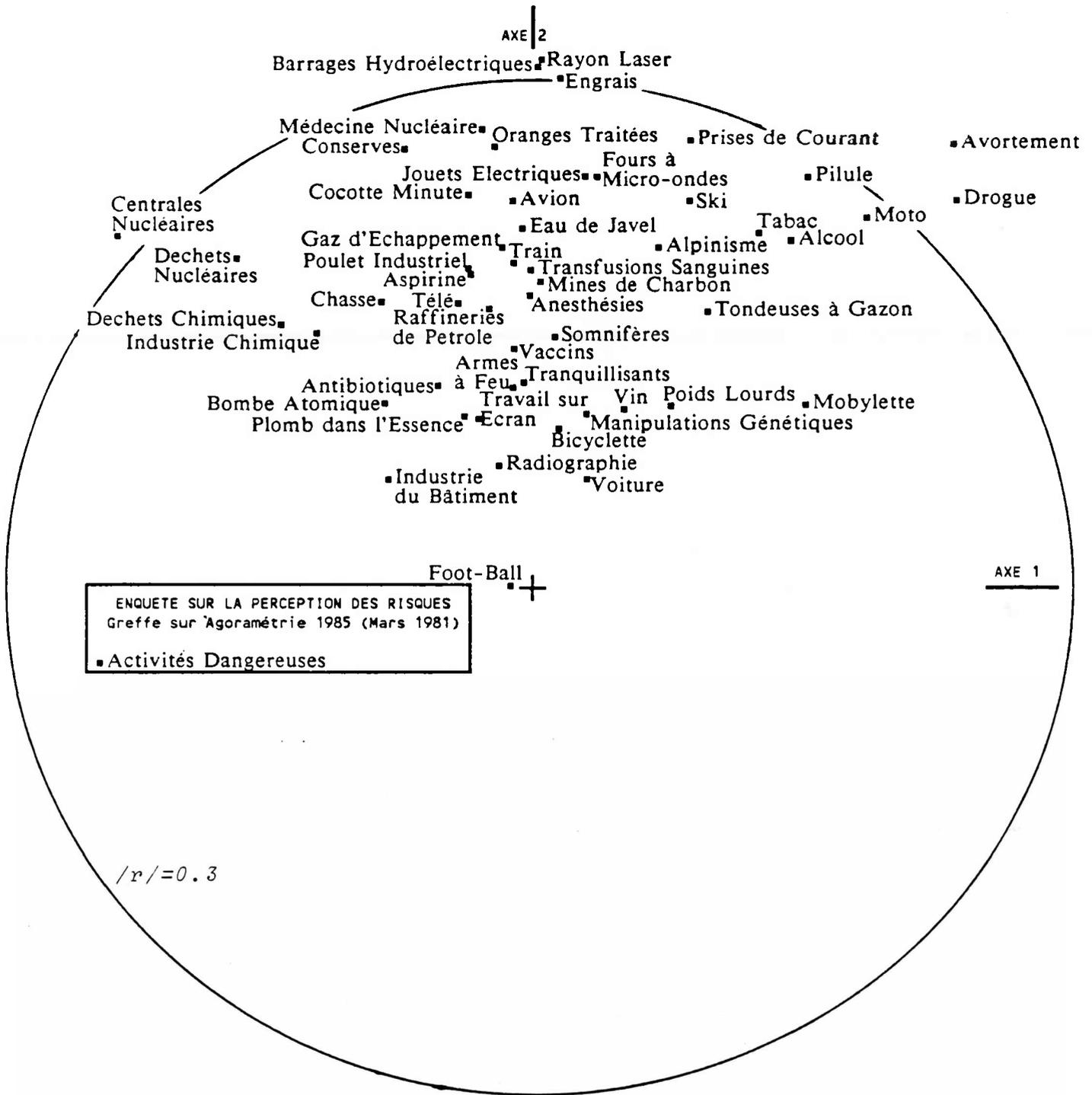
### Greffe des activités dangereuses

Dans l'enquête Agoramétrie 1985, la qualité de la reconstruction des deux indicateurs factoriels, par régression sur les variables communes, peut être considérée comme satisfaisante : les coefficients de corrélation multiple correspondants valent respectivement 0.87 et 0.92. Dans l'enquête risque, les indicateurs obtenus par la même formule de régression apparaissent comme équivalents à leurs homologues de l'enquête de référence. On ne donne pas ici les graphiques comparatifs analogues à ceux de la figure 5.6 ; ceux-ci, rappelons-le, permettent, en particulier, de contrôler la qualité de l'échantillonnage dans l'enquête particulière. Dans la mesure où les deux enquêtes ont eu lieu au même moment auprès d'échantillons représentatifs de la même population, les positions des points homologues sur les deux graphiques, après superposition, doivent être sensiblement identiques.

Les cinquante deux activités dangereuses, qui seront greffées sur le cercle des corrélations 1-2 de référence, ont été sélectionnées par un groupe d'experts. Elles sont supposées être représentatives d'un "univers" de situations à risque. Les 1000 interviewés ont donné leurs réponses sur une échelle à cinq paliers allant de "pas du tout dangereuse" à "très dangereuse".

Sur le cercle des corrélations de la figure 5.13, toutes les activités, sauf le football, se retrouvent du côté de la dramatisation : plus on dramatise, plus on a tendance à souligner le danger de certaines activités. Souvenons nous que, dans le repère de référence, la deuxième dimension agrège deux types d'attitudes : ou l'on est sur la défensive (repli) et dramatiser permet d'exprimer une angoisse qui peut trouver son origine, par exemple, dans l'isolement, ou l'on participe à une remise en cause plus idéologique (défi) et en dramatisant on exprime son opposition au libéralisme et son adhésion aux valeurs égalitaires.

Quand on juge l'avortement, la pilule, la drogue comme plus dangereuses, on est en général dans le quadrant du repli. C'est dans ce quadrant que s'expriment, en premier lieu, soit les angoisses (on reste au niveau des émotions), soit les interdits (on a recours aux stéréotypes : s'étant replié sur soi-même, on adopte des attitudes répressives vis à vis d'autrui). Quand on associe aux centrales nucléaires, aux déchets radioactifs ou chimiques, un risque plus élevé, on est en général dans



Enquête de référence : Agoramétrie 1985, n=1028 individus.  
 Enquête particulière : Risques 1986, n=1000 individus.  
 Nombre de variables de la passerelle, q=25.

FIGURE 5.13 : Greffe des activités dangereuses sur le référent Agoramétrie

le quadrant du défi où s'affirme avant tout l'égalitarisme (ceux qui se disent communistes ou écologistes se retrouvent dans ce quadrant).

Sur les cercles de corrélation des figures 5.11 et 5.13, la drogue, en tant qu'activité dangereuse se retrouve, bien entendu, dans le quadrant du repli (figure 5.13) aux antipodes du thème de conflit "le haschisch devrait être en vente libre" qui, lui, est localisé dans le quadrant de l'aventure (figure 5.11). De même, le thème "il faut continuer à construire des centrales nucléaires" (quadrant de la soumission, figure 5.11) s'oppose, comme il se doit, aux centrales réduites ici à leur facette négative d'activité dangereuse (quadrant du défi, figure 5.13). Toutes ces convergences spectaculaires viennent confirmer la bonne qualité de la greffe.

#### Vérification par l'analyse de redondance

On a vu (figure 5.13) que les dimensions 1-2 de référence, en particulier la deuxième, sont de bonnes variables illustratives dans l'analyse du risque perçu. Interprétées dans l'enquête Agoramétrie, elles présentent l'avantage considérable par rapport aux autres indicateurs du type factoriel possibles, de nous dispenser de tout un travail d'interprétation dans l'enquête risque. Mais, existe-il d'autres dimensions (en particulier, d'autres combinaisons linéaires des variables de la passerelle) présentant, a priori, une meilleure puissance explicative ? si tel est le cas : quelles sont, effectivement, ces dimensions ?

C'est pour répondre à ces questions, que l'analyse de redondance (cf 3.1.3.) des activités dangereuses par rapport aux thèmes de conflit a été effectuée.

On notera (cf. 3.1.3.)  $y^1, \dots, y^{52}$  les activités dangereuses,  $z^1, \dots, z^{25}$  les thèmes de conflit de la passerelle et  $\hat{y}^1, \dots, \hat{y}^{52}$  les parties expliquées des variables  $y^j$  par l'ensemble des variables  $z^i$  (tableau  $\hat{Y}(1000 \times 52)$ ). Pour simplifier, les variables  $y^j$  et  $z^i$  sont prises centrées-réduites. La matrice de la variance expliquée s'écrit avec les notations déjà introduites :  $R_{\hat{Y}\hat{Y}} = R_{YZ} R_{ZZ}^{-1} R_{ZY}$ . La trace de cette matrice n'est autre que la somme des carrés des coefficients de corrélation multiple associées aux régressions de chacune des 52 activités dangereu-

ses par rapport aux 25 thèmes de conflit. Ici : trace  $R_{\hat{y}\hat{y}} = 4.58$ . Pour les activités dangereuses, les thèmes de conflit n'expliquent donc qu'une petite part de la variance totale (égale à 52).

Certes, la variance expliquée est relativement faible ; il en est toujours ainsi en psychologie sociale. Mais examinons si la partie expliquée peut se réduire à un nombre moindre de dimensions. Il faut donc chercher les combinaisons linéaires orthogonales des variables  $z^i$  expliquant successivement au mieux la variance des  $\hat{y}^j$  (et donc, des  $y^j$ ). Cela s'effectue par l'analyse de redondance des  $y^j$  par rapport aux  $z^i$ , c'est-à-dire, on l'a vu, par l'ACP du triplet  $(\hat{Y}, I, D_p)$ . On est ainsi conduit à diagonaliser la matrice de la variance expliquée :  $R_{YZ}R_{ZZ}^{-1}R_{ZY}$ .

Voici les quatre premières valeurs propres de cette matrice :

Valeur propre $\lambda_i$	%	%cumulé
2.49	53.16	53.16
0.53	11.47	64.62
0.35	7.62	72.25
0.24	5.21	77.46

On voit que 53% de la variance totale est expliquée par une seule composante redondante, 65% l'étant par les deux premières composantes. Les corrélations entre composantes redondantes  $\xi^2$  et thèmes de la passerelle (variables  $z^i$ ) sont indiquées dans le tableau ci-après :

THEMES DE CONFLIT	$\xi^1$	$\xi^2$
Rétablir la peine de mort	0.4117	0.4391
Guerre mondiale inévitable	0.3144	-0.0122
Pour la patrie	0.2419	0.4553
Pour la famille	0.3278	0.5997
Maintenir la croissance	0.1320	0.3697
Montée de la violence	0.5835	0.3769
Censurer certains livres	0.5168	0.4325
Hausse des prix évitable	0.2230	-0.0564
Médecine gratuite	0.1163	-0.3356
Construire des centrales nucléaires	-0.2520	0.5743
Trop de faux chômeurs	0.4013	0.2978
Soutenir les pacifistes	0.2082	-0.2016
Ordinateurs menacent nos libertés	0.5830	-0.0172
Confiance en la Justice	-0.1474	0.2511
Utiliser l'énergie solaire	0.1414	0.0094
Moins de robots	0.7238	-0.1139
Etudiants parasites	0.3912	-0.0719
Grandes fortunes respectables	-0.3079	0.0653
La propriété c'est le vol	0.1819	-0.3660
Homosexuels comme les autres	-0.1442	-0.3013
Pollution préoccupante	0.3461	0.0224

Haschisch en vente libre	-0.3202	-0.5937
Pour l'auto-defense	0.2211	0.1479
Trop de travailleurs immigrés	0.4003	0.2958
Pas de permission de sortie	0.2884	0.3903

Les corrélations entre composantes redondantes ( $\xi^l$ ) et composantes de greffe ( $c^l$ ) sont données dans le tableau suivant :

	$\xi^1$	$\xi^2$
$c^1$	0.077	0.881
$c^2$	0.848	0.174

La première composante redondante apparaît donc très voisine de la deuxième dimension de référence (dramatisation-compromis). Corrélée positivement avec toutes les activités dangereuses (entre 0.1 et 0.4), elle semble agréger, comme  $c^2$ , les deux types d'attitude caractérisés par le repli et le défi. Notons simplement que le poids relatif de l'émotion (du côté du repli) est, ici, peut-être, plus important (corrélation élevée avec le thème "il faut diminuer l'utilisation de robots dans l'industrie" ; corrélation faible avec le thème "il faut soutenir les pacifistes").

La relation étroite qui apparaît ici entre  $\xi^2$  et  $c^1$  est un résultat remarquable. En effet, on pouvait s'attendre que la part conflictuelle de la variance des activités dangereuses soit très bien expliquée par la composante de la dramatisation. Par contre, il était nettement moins évident que la deuxième dimension redondante, celle qui permet en premier de différencier les situations à risque dans ce qu'elles ont de conflictuel, entretienne une relation aussi étroite avec cette première dimension de référence de la stabilité et du mouvement.

**Bilan : l'apport du cadre général de référence  
à travers la méthode de la greffe**

Dès que l'objet dont on parle (ici : le risque) se situe au centre des conflits sociaux, on a intérêt, et l'exemple traité le prouve avec éclat, à se replacer dans le cadre de référence des "structures de l'opinion." Bien évidemment ces structures n'expliquent pas tout ; la

variance résiduelle reste importante, la relation avec le risque étant, au delà d'une symbolique produite par le conflit, une affaire d'ajustement personnel qui implique tous les actes de la vie. Mais ce que l'on explique est une chose essentielle, car elle est à l'origine de l'émoi des responsables qui s'interrogent sur le risque perçu. On explique ici les déterminants de ce dialogue de sourds qui oppose, en général, les responsables des grandes entreprises ou administrations au public et à ses porte-parole. C'est là que peut-être se trouvent les clés d'une nouvelle forme de dialogue qui pourrait faire en sorte que les changements puissent s'opérer en souplesse. Les résultats de la greffe confortent donc la démarche constructionniste empruntée par le groupe de travail : sans être en contradiction, on est loin ici de la manière anglo-saxonne<sup>1</sup> de traiter du risque perçu.

---

<sup>1</sup> Voir, par exemple,

Slovic, P. (1987). *Perception of risk*. Science, 236, 280-285

## EPILOGUE

Tout au long de ce travail nous avons suivi un fil logique qui a conduit successivement à partir de remarques sur le phénomène de l'Opinion (*l'opinion est une force*), à un schéma (*le schéma de l'économie des rôles*), puis à un modèle (*le modèle factoriel*) devant être analysé par une technique d'analyse factorielle respectant le *principe de dualité*, puis à une procédure de recueil des données (*double tirage au hasard dans les deux univers que forment individus et variables*), puis à la mise en place d'une logistique en matière de méthodes (*greffe par régression et rotations procrustéennes*) et de traitement des données (progiciel d'analyse d'enquêtes sur micro-ordinateur), pour aboutir, enfin, à une mise à l'épreuve par des exemples de toute cette construction centrée autour de l'idée d'enquête de référence.

Le projet à long terme est bien de chercher à mettre en place un ensemble d'enquêtes, offrant des cadres généraux et complémentaires d'analyse à ceux qui oeuvrent dans le champ des sciences sociales. Dans cette entreprise ambitieuse de géomètre on en est encore aux prémices, à la faisabilité... On cherche à normaliser les manières de quantifier dans un champ qui se caractérise par une extrême diversité des concepts et des discours laissant ainsi la place à de nombreux paradoxes, voire de contradictions.

Les premiers résultats obtenus sont plus qu'encourageants...

## RÉSUMÉ

Les opinions, aussi volatiles et complexes qu'elles puissent paraître, relèvent d'un même ensemble de mécanismes. Ce postulat sous-tend l'idée d'enquête de référence : comment fournir à ceux qui oeuvrent dans le domaine de la communication sur un problème spécifique un ensemble d'outils permettant de se resituer dans un cadre général ?

Considérant que les représentations se forment, avant tout, au fur et à mesure du débat social engagé, on assimile le phénomène de l'opinion à un système qui prend en compte les ajustements qui s'opèrent entre des acteurs et le public. Toute la démarche décrite conduisant à la mise en place d'un système d'enquêtes de référence particulier découle de ce premier "schéma" : choix puis méthode d'extraction des paramètres du modèle, construction du questionnaire. C'est ainsi que l'on a été conduit à développer une méthodologie statistique et des outils informatiques qui rendent opérationnel, de façon générale, de tels systèmes d'enquêtes. La démarche et la méthodologie développées ont été mises à l'épreuve sur des exemples.

**Mots-clés :** *Opinion. Enquêtes d'opinion. Approche systémique. Analyse des données. Modèle factoriel. Analyse en composantes principales. Comparaison de structures factorielles. Régression linéaire. Rotations procrustéennes. Logiciel de dépouillement d'enquêtes.*

## ABSTRACT

Opinions, fleeting and complex though they may seem, are the concern of a common set of mechanisms. This postulate underlies the concept of a reference survey: how can people working in the field of communication be supplied with a set of tools enabling them to place their specific problems within a general framework ?

Considering that representations are building up, first and foremost, as the involved social debate goes along, we liken the opinion phenomenon to a system that takes into account the adjustments occurring between actors and the public. The considered approach that leads to setting up a particular reference survey system follows from this basic "schema": choice then extraction method of the parameters of the model, construction of the questionnaire. That is how we have been led to elaborate a statistical methodology and computer tools that make operational such survey systems, in a general way. Both approach and methodology have been put to test on some examples.

**Keywords :** *Opinion, Public opinion surveys, Systems approach, Data analysis, Factor analysis, Component analysis, Factor structure comparison, Linear regression, Procrustean rotations, Survey processing software.*

# CAHIER DE RECHERCHE

## Récemment parus :

Les opinions des Français fin 1989 et une comparaison 1981-1989, Enquête "Conditions de vie et Aspirations des Français", par Laurent Clerc, Olivier Martin, sous la direction de Georges Hatchuel, N° 11, Février 1991.

Enquête sur mémoires - Etudes et Recherches sans le secteur social : Lexicographie de la littérature grise en action sociale - Volume 1, par Séverine Binard, Michel Legros, N° 12, Mars 1991.

Construction d'un corpus et perte d'information en analyse lexicale : méthodes et pratiques, par Sébastien Lion, sous la direction de Saadi Lahlou, N° 13, Avril 1991.

Penser l'insertion - Méthodes et critères : Contribution à une analyse des critères de l'insertion dans les réseaux de prise en charge des jeunes en difficulté, par Michel Legros, N° 14, Avril 1991.

L'analyse propositionnelle du discours, par Michel Messu, N° 15, Mai 1991.

Classification dichotomique descendante, par Sébastien Lion, N° 16, Mai 1991.

Pratiques exemplaires ou exemples de pratiques : l'évaluation dans le secteur social aux Etats-Unis - Analyse de monographies présentées dans "Evaluation Review" et dans "Evaluation and the health professions", par Patricia Croutte, Michel Legros, N° 17, Juillet 1991.

Président : Bernard SCHAEFER    Directeur : Robert ROCHEFORT  
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.00

# CREDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie