

Université Paris V - René Descartes

## Rapport en vue de l'habilitation à la direction de recherches

Mars 1993

Suivi d'une postface de 2010

Philippe Cibois

### Introduction

Les 24 et 25 avril 1610, Galilée est invité à faire une démonstration de sa lunette devant 24 professeurs de l'université de Bologne. Nous ne connaissons pas le détail de ce qui s'est dit devant cette commission mais il nous reste le rapport du secrétaire qui conclut que l'instrument de Monsieur Galilée est vraiment merveilleux quand il est tourné vers un objectif terrestre, car on avait pu déchiffrer de fort loin l'inscription portée sur le fronton d'une église ; mais que tourné vers un objectif céleste, il devient fallacieux, à preuve, tourné vers l'étoile de la Vierge, celle-ci leur est apparue double<sup>1</sup>.

Il est vrai que la lunette n'était pas parfaite et que les aberrations chromatiques étaient troublantes : Képler lui-même, de bonne volonté cependant, eut bien du mal à y voir ce que Galilée y avait aperçu. Mais ce qui prouvait l'aspect trompeur de l'instrument était que l'étoile de la Vierge paraisse double. Ce qui remettait en cause l'instrument, c'était la cosmologie antique issue des pythagoriciens, revue et corrigée par Platon et qui avec la caution d'Aristote était partagée par tous. Selon cette cosmologie, l'univers est divisé en deux : il y a le ciel qui est l'âme du monde, où donc l'étoile de la Vierge est son âme, qui ne saurait d'aucune façon être double, et il y a le monde que nous connaissons, le monde sublunaire, soumis au changement et à la corruption. La lunette fonctionne pour notre monde à nous et non pour le Ciel : cette opposition qui nous semble aujourd'hui non pertinente était encore très forte à l'époque.

Elle ne le resta pas longtemps, l'observation du relief de la lune et des satellites de Jupiter put se faire bientôt avec des instruments améliorés et très rapidement le monde scientifique, y compris les jésuites de l'observatoire pontifical, admirent ce fondement de la science moderne qu'il n'y pas deux physiques, celle du ciel et celle de la terre mais qu'il n'y en a qu'une, celle qui s'impose à nous tous aujourd'hui.

Si j'éprouve le besoin d'ouvrir ce rapport d'habilitation par cette histoire, c'est à un double titre : pour montrer l'importance capitale qu'a eue la vision instrumentale au moment de la naissance de la science moderne, et sur ce point de la vision instrumentée par le graphique je reviendrai longuement ; et pour nous contraindre à plus de modestie. En effet, il nous est facile à bientôt quatre siècles de distance, de nous moquer de ces pauvres universitaires de Bologne qui avaient des yeux pour ne pas voir : il y aura peut-être dans quatre autres siècles un archéologue de la science de la fin du XXe siècle qui aura à notre égard le même sourire condescendant et qui dira que nous avons de magnifiques

---

<sup>1</sup> Cité dans Feyerabend 1979 : 133

recensements, des enquêtes innombrables, des instruments d'observation puissants et raffinés mais que ce qui nous empêchait de comprendre le social, c'était l'idée que nous nous en faisons.

On peut réagir à cette mise en garde en éprouvant de la méfiance face à celui qui la fait en se demandant s'il ne se prend pas pour un nouveau Galilée, fondateur d'une nouvelle science du social grâce à une instrumentation puissante, des techniques statistiques nouvelles, des logiciels sophistiqués. Je dois dire immédiatement que ce rapport d'habilitation montrera comment une telle tentation peut venir, comment elle peut avoir des conséquences positives, comment on peut chercher à en accepter certains aspects et en rejeter d'autres, comment on peut chercher à réfléchir sur une pratique moins ambitieuse mais plus conforme à la tradition sociologique.

\*\*\*

## Chapitre 1 : les éblouissements

### Le Musée des Arts et Traditions Populaires

En 1971 le musée des Arts et Traditions Populaires brillait de tous les feux de la modernité : un bâtiment neuf dans un site prestigieux, dans lequel on s'apprêtait à ouvrir une exposition permanente et une galerie d'études selon des principes muséographiques mis au point par G. Rivière et qui associaient un objet traditionnel et une méthode scientifique. En un mot le paradis sur terre pour un étudiant qui en franchissait les portes avec émotion pour venir y commencer un stage dans le service informatique. Car ce jeune musée disposait de ce sacrement de la scientificité qu'était l'ordinateur, sacrement car il en était signe efficace. Symbole d'abord, prise de position de la nécessité d'une étude formelle et structurale des collections d'objets domestiques, des proverbes ; mais aussi mise en œuvre efficace de ce projet scientifique puisqu'il permettait l'étude concrète de ces objets avec les méthodes nouvelles de la classification automatique et de l'analyse des correspondances.

Car j'y venais pour faire un stage d'informatique, pour appliquer sur des vraies données les techniques que j'étais en train d'apprendre dans un certificat d'Analyse des données. Cet enseignement m'avait également permis d'avoir mon premier contact avec l'informatique lors d'un stage d'initiation dans le service de calcul de la Maison des Sciences de l'Homme.

### La Maison des Sciences de l'Homme

La MSH, comme les ATP, disposait d'un service informatique équipé d'ailleurs du même ordinateur, un 1130 IBM doté de 8K de mémoire à la MSH et de 2K aux ATP. Les raisons en étaient les mêmes : dans des bâtiments neufs, dans des structures neuves, il fallait montrer que la Science de l'Homme avait partie liée avec la recherche la plus scientifique qui soit, celle qui utilise les techniques les plus avancées. J'étais tout à fait mûr pour reprendre à mon compte cette manière de voir : j'avais, enfant, été fasciné par le monde de la technique et de la recherche scientifique qui lui était intimement associée. Les expéditions polaires de P.E. Victor, la vulcanologie d'H. Tazieff, la conquête de l'espace avaient nourri mon imaginaire et un théodolite me paraissait, par l'arpentage scientifique qu'il permettait, le plus bel instrument de recherche du monde.

Cependant, comme ma vocation géologique avait été contrecarrée par une vocation religieuse, et quand cette dernière se révéla assez rapidement pour mon compte une impasse, une synthèse des deux se fit dans la lecture de Durkheim où je découvris avec passion qu'il était possible de considérer les phénomènes humains comme des choses que l'on pouvait étudier d'une manière scientifique. En ce mois de mai 1968, tandis que j'exhortais les fidèles

dont j'avais alors la charge à comprendre que l'Esprit souffle où il veut, je découvrais dans *Le suicide* que ce plan d'action pouvait être mis en pratique d'une manière convaincante.

En descendant donc au deuxième sous-sol de la MSH pour y apprendre le Fortran, j'eus le sentiment très vif de participer à une expérience à la fois nouvelle, la découverte scientifique de la réalité sociale, et ancienne, ce regard scientifique sur le monde qui du mythe pythagorique aux expériences scientifiques d'Archimède s'ancrait dans une antiquité aussi respectable que celle de la tradition judéo-chrétienne et qui semblait l'assumer.

## Une enquête sur le Patrimoine

Le Fortran par sa rudesse, se révéla une école de rigueur : qu'il faille, pour afficher un résultat numérique, prévoir le nombre de blancs à sauter, la largeur du libellé à inscrire et tenir compte de la forme numérique du résultat me fit prendre conscience immédiatement que l'ordinateur n'était pas un être mythique doté de tous les pouvoirs mais une bête machine à laquelle il fallait mâcher le travail. Je me mis à la programmation qui emplit rapidement mes jours, et aussi mes nuits comme pour tout programmeur débutant et inexpérimenté. En effet pour les besoins du stage aux ATP je devais procéder à la programmation d'outils de dépouillement d'enquête : tri à plat, tri croisé, recodage, mise en codage disjonctif. Je mis là au point des morceaux de programmation dont certains me servent encore. J'éprouvais du plaisir à la chose car j'y trouvais une synthèse de la science et de l'étude des phénomènes humains.

Cette étude se situait à un niveau modeste : il s'agissait de réexploiter une enquête sur le patrimoine qui n'avait été exploitée que sous forme de tris croisés. Inutile de dire que pour le milieu de l'Analyse des données où se faisait l'enseignement et le stage, il s'agissait là d'une sous-exploitation grave qu'il fallait pallier par des analyses factorielles. Ce milieu de l'Analyse des données dans lequel se trouvaient des anciens élèves de J.P. Benzécri : B. Leroux, G. Oppenheim et Cl. Deniau pour l'enseignement, M. de Virville pour le stage, vivait dans l'enthousiasme de la découverte que l'analyse factorielle des correspondances, récemment mise au point, permettait de prendre en compte en même temps plusieurs questions d'une enquête afin d'en avoir une image globale.

Appliquée à l'enquête sur le patrimoine, je vis apparaître que l'idée que l'on se fait du rôle du patrimoine dépend de sa possession. Que les cadres supérieurs, qui en ont un, y croyait alors que les cadres moyens qui n'en n'ont pas n'y voyait qu'un principe moral, des valeurs qu'il faut transmettre. Cela me paraissait fort logique et je n'eus pas le sentiment d'avoir fait avancer de beaucoup la connaissance scientifique des réalités humaines : ce ne fut pas l'avis de Michel de Virville qui, insistant sur l'aspect méthodologique de l'utilisation de l'analyse factorielle, me proposa immédiatement d'exposer mes résultats à un colloque scientifique.

Ce colloque, qui eut lieu au couvent de l'Arbresle<sup>2</sup> et qui dans l'esprit d'un certain nombre de ses participants devait permettre la confrontation des techniques formelles avec les sciences de l'homme (c'est ce qu'espérait par exemple le créateur de l'analyse de similitude Cl. Flament), se révéla entre les mains de son organisateur, J.P. Benzécri, un instrument à la gloire de l'analyse des correspondances et de la classification automatique. Je présentais mon plan factoriel (mal, j'avais oublié de représenter les axes : Cibois 1971), Benzécri me posa quelques questions techniques et s'extasia sur la puissance de la méthode. Ce fut ma première rencontre avec lui, personnage déjà mythique à la longue barbe, au béret et aux bottes de caoutchouc, drapé, quelle que soit la température dans un vaste vêtement à la couleur

---

<sup>2</sup> Colloque de l'Arbresle sur l'Analyse des données : 21-22 avril 1971

indéfinissable mais qui d'une voix chaude et claire savait poser les questions qui montraient son intérêt profond pour les données qui étaient traitées devant lui.

Ce colloque fut le premier de ceux qu'il organisa à la gloire de ses techniques d'analyse de données : il me valut d'être cité dans l'index des noms des deux tomes de l'Analyse des données (Benzécri 1973) pour être dans la même publication qu'une étude de M. de Virville, un des soixante-dix<sup>3</sup> collaborateurs de Benzécri.

A la fin de l'année, j'étais bien formé au dépouillement d'enquête en utilisant l'analyse factorielle mais je ne voyais pas trop ce que cela représentait car pour ce qui est de la méthode elle-même, je savais la faire à la main sur de petites données en utilisant règle à calcul et machine mécanique à additionner (bonnes aujourd'hui pour les ATP), mais je fus noyé dans l'algèbre linéaire qui ne m'apporta pas beaucoup de lumières sur la technique. J'avais beau ne pas être bloqué face aux mathématiques qui me semblaient le support de toute science, la barre avait été mise un peu haute. Il n'empêche, notre groupe d'étudiant essayait les plâtres d'un nouvel enseignement et il le faisait avec des enseignants à l'enthousiasme communicatif : des solidarités sont nées qui subsistent encore.

## L'analyse empirique de la causalité

Le cours de R. Boudon l'année suivante fut le premier cours de sociologie qui me parut répondre au programme durkheimien que j'avais adopté : auparavant je n'avais entendu qu'un althussérisme dopé par le contexte d'après 68 puis une sociologie générale où l'on discutait des avantages et des inconvénients de considérer la société comme un système. J'avais eu dans mes études antérieures mon compte de théologie, j'avais bien l'impression de retrouver dans ces deux cas la même attitude et je me fermais attentivement les oreilles et ne laissais filtrer que le minimum nécessaire à la passation des examens.

Avec la découverte de la sociologie empirique introduite en France par les ouvrages de Boudon et Lazarsfeld<sup>4</sup>, il me semblait être au cœur du sujet : j'avais sous les yeux une sociologie qui prenait les données au sérieux mais qui me posait quand même quelques problèmes. Certes R. Boudon n'utilisait pas l'Analyse des données mais il me semblait normal qu'il l'ignorât : il suffisait de lui en montrer les mérites et c'est pourquoi je voulu en faire une démonstration dans mon mémoire de maîtrise passé avec lui.

Mais plus profondément, ce qui me posait problème c'était l'usage constant qui était fait de concepts qui me semblaient typiquement être des prénotions durkheimiennes : quand on se posait le problème de savoir si la mobilité sociale avait cru ou décru, il me semblait que les indices les plus sophistiqués qui tenaient compte de la mobilité structurale n'étaient quand même que l'opérationnalisation de l'idéologie nord-américaine de la société ouverte. De même quand R. Boudon nous présentait ce qui allait devenir *L'égalité des chances* (Boudon 1973), je ne pouvais m'empêcher de penser que savoir si l'éducation s'est démocratisée ou non est foncièrement un problème politique et social, non un problème sociologique ni a fortiori scientifique, quelle que soit la sophistication des méthodes employées.

J'avais eu auparavant ma dose de "social" et bien sûr de "religieux" dans les institutions ecclésiastiques et je n'avais nulle envie de me replonger dans des problèmes de cet ordre, même avec le regard armé du sociologue. C'est pourquoi après un court passage auprès d'H. Desroche, qui m'appriis ce qu'est le travail intellectuel (Desroche 1971), je renonçais à me spécialiser en sociologie de la religion où évidemment j'avais un terrain tout trouvé, des introductions, une problématique, une compétence. Je refusais même une spécialisation sur un

---

<sup>3</sup> Allusion voulue de Benzécri à la traduction grecque de la Bible due aux Septantes

<sup>4</sup> *Le vocabulaire des sciences sociales, 1965* et *L'analyse empirique de la causalité 1966*

terrain quelconque et décidais de continuer dans ce qui m'avais séduit, c'est à dire le terrain des méthodes d'analyses de données, pour voir en quoi elles pouvaient être utiles au sociologue.

## Etude d'une cohorte d'élèves

Mon mémoire de maîtrise fut un premier galop d'essai : par relation, je récupérai des données non traitées sur une cohorte d'élèves saisis en fin de troisième et suivis pendant trois ans. Je sortis des services statistiques du Ministère à Vanves avec un paquet de cartes perforées sous le bras et une totale liberté. Comme d'autre part R. Boudon était plutôt un directeur de maîtrise libéral, je n'eus pas beaucoup de contraintes et pu faire fonctionner à ma guise les techniques d'Analyse de données : je fréquentais assidument le centre de calcul du CNRS à Orsay où la carte perforée était alors reine. L'IBM 360 était alors utilisé et j'appris une syntaxe de cartes de contrôle qui allait me servir pendant environ 15 ans jusqu'à l'arrivée de la micro-informatique.

Utiliser la carte perforée pour des fins de recherche sociale n'était d'ailleurs qu'un juste retour des choses, car cet instrument était né à la fin du XIXe siècle aux Etats-Unis pour les besoins du recensement : celui de 1890 fut traité très rapidement grâce à la mécanographie de Mr Hollerith dont l'invention fut exportée et qui fonda dans les années 20 une *International Business Machine Corp.* qui sous les initiales d'IBM allait connaître le développement que l'on sait. La carte perforée fut antérieure à l'informatique, elle lui survivra sous la forme de cette largeur de 80 qui subsiste aujourd'hui encore pour les écrans et qui marquera peut-être encore le siècle suivant, vestige d'un usage ancien comme le sont les revers de nos vestes.

Le problème de la cohorte n'était pas trop compliqué : il y avait des variables de statut social et des variables de comportement. Il suffisait d'appliquer la technique alors récente des variables supplémentaires en analyse factorielle. Mais je commençais à me poser un problème qui allait m'occuper de nombreuses fois dans la suite : quel était le statut du plan factoriel obtenu. On y voyait certes que les modalités relatives aux filières nobles étaient proches des modalités élevées de statut social. De ce point de vue tout était compréhensible, mais quel était la réalité de cette proximité : pour y voir plus clair je commençais par simplifier au maximum l'analyse pour ne plus avoir comme variables actives que les situations successives dans le temps. En projetant en même temps les numéros des individus, je vis alors que les configurations de trois situations (en seconde, en première, en terminale) des diverses filières du système correspondait à un nombre limité de cas de figure. En faisant l'inventaire des individus qui correspondait à ces cas de figure je m'aperçus que j'étais en train de réinventer le tri de profondeur trois. Pour chacune des modalités des trois questions, il s'agissait de faire un inventaire lexicographique de toutes les situations et de compter ensuite les individus. Pour revenir aux données à partir de l'analyse factorielle, il suffisait donc, pour chacune des situations répertoriées, de croiser la sous-population obtenue avec les variables de statut. On obtenait ainsi facilement les relations entre variables.

## Le retour aux données

En faisant cette recherche, je mis à jour une démarche dont je m'aperçois aujourd'hui qu'elle continue à me guider dans ma manière d'utiliser l'analyse factorielle. En effet, en traitant une enquête par l'analyse des correspondances, on voit qu'un plan factoriel montre des proximités entre des modalités de réponses issues de plusieurs questions. Quelle est la nature de ces proximités, quel est leur caractère de généralité ? Quand on voit une proximité entre filière défavorisée et situation sociale inférieure, doit-on dire que l'analyse des correspondances montre que les défavorisés sont exclus ? Ou plus encore le prouve ?

Une telle manière de penser est très spontanée : en effet, le chercheur qui travaille dans le champ de la sociologie de l'éducation a déjà maintes et maintes fois vérifié qu'il y avait une liaison entre des situations sociales et des filières de scolarisation : comment se fait-il qu'en voyant un tableau croisé où cette liaison apparaît une nouvelle fois, il a l'impression d'une redondance tandis que s'il voit un plan factoriel qui lui dit la même chose, il a l'impression qu'une "preuve" nouvelle est apportée ? On verra plus loin la nature de cette liaison, pourquoi le chercheur a le sentiment de preuve (illusoire) et comment il peut éventuellement trouver une preuve réelle par une démarche post-factorielle. A cette époque je résolus empiriquement le problème par une utilisation des techniques graphiques de Jacques Bertin.

## Sémiologie graphique

Ce livre était récent (Bertin 1967). Il rassemblait les résultats obtenus au Laboratoire de graphique de la 6e section de l'EPHE de l'époque. J. Bertin y montrait brillamment que la perception visuelle a un pouvoir synthétique d'une puissance telle que des résultats numériques doivent être impérativement traduits en mode graphique si on veut dépasser l'analyse locale. Comme je venais de découvrir l'analyse des correspondances et son mode d'exposition sous forme de plan factoriel, je ne pouvais qu'être séduit par cette prise de position qui était d'ailleurs rendue inattaquable par une multitude d'exemples issus de recherches réelles. Mais une autre conclusion de J. Bertin était que si tout graphique est porteur d'information, celle-ci peut être fallacieuse. Que par exemple sur le graphique des cohortes, pour rendre compte de l'importance numérique des diverses filières, il fallait tracer des traits dont l'épaisseur devait être proportionnelle non pas aux effectifs mais à leur logarithme.

Ce livre montrait que la vision a un rôle fondamental dans l'expression de résultats numériques mais qu'elle a ses lois propres : les ignorer pouvait être dangereux. Il y eut immédiatement dans mon esprit deux catégories d'individus : ceux qui avaient compris le projet de Bertin et avec qui on pouvait discuter et les autres, car j'avais l'impression que ceux qui acceptaient le passage par une expression graphique étaient ceux qui utilisaient des chiffres, c'est à dire des méthodes scientifiques. Quant aux autres, leur refus des chiffres les renvoyait dans l'interprétation gratuite. Que toute idée, si elle ne se confronte pas d'une manière ou d'une autre avec le réel, ne peut être invalidée et donc, en suivant les règles poppériennes apprises, ne pouvait être acceptée comme prouvée.

Dans la priorité donnée aux expressions graphiques, il y avait plus qu'un hommage mérité à J. Bertin, il y avait un choix d'une vraie méthode scientifique appliquée à la réalité sociale. Il y avait le parti pris de dire qu'il n'y a pas deux sciences, une de la matière et une des phénomènes sociaux. Si l'on croyait à la science de la matière, il fallait appliquer ses méthodes aux réalités humaines : cela ne pouvait passer que par des confrontations empiriques falsifiables et donc exprimables graphiquement. L'utilisation du graphique devenait ainsi pour moi la pierre de touche de la scientificité.

J'étais donc armé de certitudes et de techniques : R. Boudon me proposa d'être son collaborateur technique dans son équipe CNRS du GEMAS<sup>5</sup>. Je commençais par vérifier les chiffres de *L'inégalité des chances* : je mis au point un programme qui permettait de plus de faire varier les paramètres des équations. Les autres travaux concernant la sociologie de l'éducation, ce ne furent plus des travaux de débutant. En même temps, Marc Barbut me proposa de faire des travaux de programmation pour des historiens et des archéologues.

---

<sup>5</sup> Groupe d'Etude des Méthodes de l'Analyse Sociologique

## Premiers travaux informatiques

Il s'agissait de répondre à la demande précise d'un historien, R. Mousnier qui pensait avoir un problème qui ne pouvait trouver une solution que dans l'informatique. Si cette question n'est pas directement pertinente pour la sociologie, elle l'est pour l'organisation de la recherche utilisant l'informatique et l'expérience acquise sur ce terrain fut utile pour l'application de l'analyse des données à la sociologie.

Le problème semblait au chercheur gigantesque : il avait une nomenclature d'un millier de titres et il cherchait à voir qui se mariait avec qui. Il avait donc besoin d'un tableau de 1000 sur 1000 pour entrer ses données : il alla voir les gens d'IBM et leur posa le problème. Un ingénieur lui répondit qu'un tel tableau ne pouvait tenir que dans les plus gros ordinateurs de l'époque. Il fit un devis que le chercheur transmit à son président d'université qui déclara forfait immédiatement. Il alla pleurer misère chez l'EPHE sa voisine : et après quelques péripéties M. Barbut me confia le dossier.

Spontanément je proposai de réduire les catégories de 1000 à 100 en arguant de l'exemple du code des catégories socio-professionnelles : ce ne fut pas considéré comme une offense au vu de ma bonne volonté évidente mais on m'expliqua qu'une société où la hiérarchie est basée sur le métier peut bien avoir un code réduit de profession, tandis qu'une société d'ancien régime, qui était une société d'ordre, ne pouvait admettre pareille réduction.

L'examen des données traitées ramena le problème à ses justes proportions : il y avait peut-être un million de cas possibles mais il y avait moins d'un millier de mariages à étudier. Il suffisait donc de mettre en ordre lexicographique les dossiers en triant d'abord sur le titre du mari, puis à égalité de ce point de vue sur le titre du père de la mariée. Le problème fut vite réglé à la grande satisfaction du chercheur.

La première conclusion méthodologique que j'en tirai fut que pour pouvoir utiliser correctement l'informatique et les méthodes quantitatives, il faut connaître à la fois la technique et le problème à traiter. Il faut donc accepter de passer beaucoup de temps pour entrer dans la problématique du chercheur : le chercheur lui, ne fait pas la démarche inverse puisqu'il a sous la main quelqu'un de compétent. On a là la limite du travail "à façon" en informatique : pour travailler correctement, il faut que le spécialiste qui dispose de la technicité passe beaucoup de temps à comprendre la problématique du chercheur qui lui n'a pas d'acquisition à faire. Evidemment, les spécialistes traités de cette façon, devant l'asymétrie des investissements ont le sentiment de se faire exploiter et se refusent à "perdre du temps" à écouter le problème du chercheur. La situation redevient symétrique : plus personne ne fait d'effort pour comprendre l'autre et le résultat peut être prévu à l'avance : il sera insatisfaisant pour les deux parties. Le spécialiste aura le sentiment que des méthodes géniales sont mal utilisées par le chercheur et le chercheur pensera que le spécialiste n'a à lui proposer que des méthodes polyvalentes et passe-partout qui ne s'appliquent évidemment pas à son cas, si ce n'est qu'au prix de réductions qu'il juge intolérables.

Un deuxième travail me fut proposé en collaboration avec Jean-Pierre Bardet qui exploitait les rôles fiscaux de Rouen au XVIIe et XVIIIe siècles. Comme j'avais tiré les leçons de l'expérience précédente, je commençais par me faire expliquer en profondeur, le problème à traiter (faire des statistiques et les cartographier), la nature des données et des codages déjà effectués puisque les données avaient déjà été encodées. Ce fut ce dernier point qui me posa d'ailleurs le plus de problèmes : il fallut pratiquement faire un programme de reconnaissance des formes pour identifier tous les cas plus ou moins complexes qui avaient été prévus et créer une base de donnée exploitable facilement dans la suite par des programmes créés pour la

circonstance (si les Systèmes de gestion de bases de données existaient déjà, il n'étaient pas encore disponibles sur le CIRCE à Orsay).

Grâce à une bonne volonté réciproque, le traitement arriva à ses fins (Bardet 1983) mais du fait de la taille de l'investissement en temps qui nous fut nécessaire, J.P. Bardet décida ensuite d'apprendre la programmation afin de retrouver la liberté d'exécution qu'un chercheur juge normale. Pour ma part, j'en tirai ma deuxième conclusion méthodologique : plutôt que de forcer un chercheur à passer du temps (à fonds perdu) pour expliquer son problème à un spécialiste, il valait mieux lui faire gagner du temps en lui apprenant la programmation. L'apprentissage est rentable pour le chercheur puisqu'il devient autonome et peut multiplier ses demandes. Il peut même infléchir sa recherche en fonction des premiers résultats, ce qui semble la démarche normale du chercheur. Au contraire quand le travail à façon est fait dans sa manière traditionnelle, le spécialiste commence par demander l'intégralité des travaux à effectuer avant de faire son estimation du cout nécessaire (en temps, en général), qui ne peut plus être modifiée ensuite au grand désespoir du chercheur.

## Informatique et Sciences Humaines

Je commençais à bien réfléchir sur ces problèmes de méthodologie de l'informatique en sciences humaines et de ce fait Marc Barbut m'offrit de participer au comité de rédaction de la revue *Informatique et Sciences Humaines* et il me proposa de faire une enquête pour le compte de la DGRST (Cibois 1973) sur l'utilisation de l'informatique dans les sciences de l'homme.

Je circulai dans une cinquantaines d'équipes de recherches utilisant l'informatique en sciences humaines et j'interviewai les chercheurs sur les problèmes rencontrés : j'y vérifiai ce que j'avais déjà vécu pour mon compte, c'est à dire que les résultats sont bons quand il y a un interface de connaissances entre le chercheur et l'informaticien. A contrario je produisis de nombreux exemples de mauvais résultats quand cet interface n'existait pas. Au comité de rédaction d'*Informatique et Sciences Humaines* ce rapport fut apprécié, en particulier par Henry Rouanet avec qui ce fut mon premier contact.

Pour les besoins de la revue, j'entrai en contact avec divers milieux utilisant l'informatique et en particulier avec un laboratoire qui investissait beaucoup dans le domaine de la formalisation, le CADA<sup>6</sup> à Marseille. Je fus ainsi amené à assister au colloque de 1972 sur les bases de données archéologiques consacré aux amphores.

Le problème de la formalisation des données relatives aux amphores est intéressant du point de vue de la formalisation en sciences de l'homme en général. A l'origine les amphores sont de simples récipients, la boîte de conserve de l'antiquité, mais ce sont pour nous de beaux objets, recherchés comme tels. Cependant la vision des archéologues étant passé au fil des siècles de la recherche des beaux restes de l'antiquité à l'anthropologie et à ses méthodes, les restes d'amphores servent aujourd'hui à repérer les flux économiques.

Pour les besoins du CIL, *Corpus des inscriptions latines*, vaste compilation d'inscriptions commencée au XIXe siècle, un chercheur, Dressel, mis au point une typologie des formes d'amphores afin de ne pas avoir à décrire le détail de chaque récipient sur lequel se trouvait une inscription. On peut ainsi parler des "Dressel 2-4", amphores classiques, hautes, ou des "Dressel 20", amphores ventrues. Cette typologie purement visuelle était considérée comme insatisfaisante et des archéologues collaborèrent avec l'équipe de Jean-Claude Gardin

---

<sup>6</sup> Centre d'Analyse Documentaire pour l'Archéologie

qui avait introduit le souci de la formalisation documentaire, puis de la formalisation des objets en archéologie<sup>7</sup>.

Les profils d'amphores furent d'abord saisis d'une manière exhaustive en notant la coordonnée de chaque point de la courbe. Chaque amphore était décrite avec une extrême précision mais il apparut vite que cette masse de chiffres, si elle autorisait un stockage et une reproduction parfaite, ne permettait ni de retrouver les types, ni de les améliorer. Des algorithmes performants ne sortaient rien de ces données, noyés qu'ils étaient sous la masse des chiffres non pertinents.

Pour sortir quelque chose d'intéressant de ces données, il fallait trouver les traits pertinents, sortir de la perspective "étic" (de l'anglais phonetic) pour passer à la perspective "émic" (de phonemic) et donc passer, analogiquement, de la phonétique à la phonologie ; de la description physique, complète, gigantesque, à la recherche des traits pertinents, ceux qui manifestent les oppositions sur lesquelles viennent se greffer l'effet de sens. Cette manière de faire, inspirée du structuralisme linguistique, donna de bons résultats sur les amphores : il suffisait en effet de considérer quelques rapports de forme comme celui du périmètre maximum sur la hauteur. On voyait que certains types correspondaient à certaines plages de ce rapport et que combiné avec la présence ou l'absence de certains traits on retrouvait et précisait la classification de Dressel.

A travers cet exemple je fus convaincu de la nécessité de propager à d'autres secteurs des sciences humaines, et en particulier en sociologie, ce paradigme linguistique, et l'analyse factorielle me sembla un bon moyen pour y arriver dans la mesure où la perspective "(phon)étique" conduisait à une description précise mais exagérée, redondante, dont une méthode descriptive ample d'analyse des données pourrait venir à bout et suggérer au chercheur une perspective "(phon)émique-(phon)ologique" qui lui permettrait de découvrir les oppositions pertinentes.

J'avais la méthode, j'avais le paradigme, la quête du Graal pouvait commencer et je n'étais pas près d'oublier la leçon de J.-Cl.Gardin. L'article que j'en tirai pour *Informatique et Sciences Humaines* (Cibois 1975) fut remarqué par Mario Borillo, un collaborateur de J.-Cl.Gardin, qui me proposa de participer à la réorganisation du Service de Calcul de la MSH : ce fut l'expérience du LISH, *Laboratoire d'Informatique pour les Sciences de l'Homme*.

## L'approximation

A cette époque (1975), j'assistai à une session de formation du CNRS "Mathématiques et sciences humaines", organisée par le CMS, Centre de Mathématiques Sociales de l'EHESS, et animée par Georges Guilbaud.

G. Guilbaud, appelé par Cl. Lévy-Strauss et F. Braudel à créer au sein de la 6e section de l'EPHE un pôle de mathématiques qui reprendrait la mathématique sociale de Condorcet, avait créé le *Centre de Mathématiques Sociales* qui, avec la participation de Marc Barbut et de Jacques Maitre, fut dans les années soixante et au-delà un centre actif de diffusion des techniques mathématiques, en particulier en sociologie au sein du Centre d'études sociologiques de la rue Cardinet.

Mathématicien de J. Lacan (Roudinesco 1986 : II, 564-567), créateur de films mathématiques, G. Guilbaud m'a fait comprendre ce qu'étaient les mathématiques et également ce que doit être un cours, ce mélange de séduction, de mise en scène, d'attention aux manières de voir des auditeurs, cette simplification extrême du problème quand on le

---

<sup>7</sup> Bibliographie in Gardin 1991

domine, ce passage par le calcul qui évite le formalisme mathématique. Il revisita devant nous l'analyse factorielle en la replaçant dans la logique qui est la sienne, celle de la gestion correcte de l'approximation, de l'à-peu-près (Guilbaud 1985).

En effet, il suffisait de considérer le tableau dont on voulait faire l'analyse factorielle (et le principe vaut aussi bien pour l'analyse en composantes principales que pour les correspondances), comme un tableau à décomposer en une somme de tableaux simples, connus par leurs marges (de rang un) mais dont le premier soit la meilleure approximation possible. En correspondances, c'est déjà le cas avec le tableau des effectifs théoriques sous l'hypothèse d'indépendance entre lignes et colonnes. On peut dire que le tableau se décompose en indépendance (de rang un) et écarts à l'indépendance. On recommence la décomposition avec les écarts à l'indépendance dont on trouve une approximation de rang un grâce à un algorithme des puissances itérées, tout à fait faisable à la main. On extrait ainsi un résumé ligne et colonne qui par multiplication donne la première approximation des écarts, et ainsi de suite jusqu'à épuisement du rang de la matrice de départ.

Ce sont ensuite ces résumés, ces facteurs, qui permettent à la fois la reconstruction d'une matrice d'approximation et une représentation factorielle. Si un point ligne et un point colonne sont en proximité angulaire (en conjonction pour reprendre le langage des astrologues très cohérent ici, cf. Cibois 1993), cela manifeste que dans les premières approximations il y a un écart positif à l'indépendance et donc une liaison. Si c'est une quadrature, il y a indépendance, si c'est une opposition, un déficit.

Avec quelques concepts simples, bien connus déjà des utilisateurs de sciences humaines comme la matrice de rang un analogue à l'indépendance ou la notion de facteur comme résumé d'un tableau de rang un, G. Guilbaud venait de reléguer le formalisme géométrique de l'algèbre linéaire au rang des outils complexes, indispensables au mathématicien mais inutilisables pour le commun des praticiens. Il lui substituait un corps de concepts accessibles à tout un chacun qui voudrait faire l'effort de se l'approprier. Non seulement j'avais une bonne méthode mais je disposais maintenant d'un mode d'exposition génial pour la faire comprendre.

Le séminaire de G. Guilbaud devint pour moi un endroit de rêve où j'appris comment on fait un calendrier, et en particulier pourquoi il y a des hésitations dans le calendrier arabe, comment on compte les vis et les boulons en se servant des logarithmes, comment est faite la gamme, en quoi consistent les rapports de Pythagore et de l'harmonie, en quel sens Archimède connaissait les logarithmes ou au moins le travail sur les exposants : ce que l'on découvre en lisant l'Arénaire, ce que je fis avec attention et où je découvris l'astronomie antique et comment on faisait des expériences de physique dans l'antiquité<sup>8</sup>.

\*\*\*

J.-P. Benzécri et l'analyse des correspondances, J. Bertin et le sémiologie graphique, R. Boudon et l'analyse empirique en sociologie, J.-Cl. Gardin et le structuralisme linguistique en sciences de l'homme, G. Guilbaud et les mathématiques, furent les gens qui me formèrent et m'éblouirent.

\*\*\*

---

<sup>8</sup> in Tome II des œuvres d'Archimède de l'association Guillaume Budé, Pars, les Belles Lettres, 1971

## Chapitre 2 : sociologie de l'éducation

### L'enseignement supérieur court

Mon premier travail de sociologue professionnel se fit au GEMAS sous la direction de Janina Lagneau : il s'agissait d'un contrat OCDE qui consistait à comparer les enseignements supérieurs courts de trois pays de l'OCDE : France (IUT), Grand- Bretagne (Polytechnics) et Yougoslavie (Visa Skola). Il me permit de comparer les résultats que J. Lagneau avait produits pour son travail sur les IUT (Lagneau 1973), avec des données analogues élaborées en Grande-Bretagne et en Yougoslavie. Les difficultés furent nombreuses : le questionnaire anglais était différent, mais les données étaient encodées sur une bande standard. Si pour la Yougoslavie le questionnaire était le même que pour les IUT, le codage informatique était assez exotique. Je fis ainsi l'apprentissage que le plus difficile en analyse secondaire n'est pas seulement de se procurer les données, car les institutions sont plutôt fières qu'on veuille utiliser les données qu'elles ont produites, le plus dur, c'est de les lire car, du fait de la division du travail entre sociologue et informaticien, les descriptions ne correspondent pas toujours à la réalité physique de ce qui est enregistré. Il faut se livrer à un travail de détective, recouper les résultats obtenus avec les résultats déjà publiés, en un mot faire des hypothèses de codage avant de faire des hypothèses de recherche.

Voulant absolument travailler sur les données d'origine et non sur des tableaux croisés faits auparavant, j'arrivai à mes fins et pu ainsi mener une analyse comparative sur des bases bien assurées. Je pu ainsi montrer que contrairement à la stratégie française de l'époque en matière d'enseignement supérieur court, il était assez contradictoire de vouloir augmenter le prestige des IUT pour favoriser leur développement. Les expériences anglaises des Polytechnics et Yougoslaves des Ecoles supérieures montraient que l'enseignement supérieur court se développe quand il n'est pas choisi pour lui-même mais faute d'avoir pu entrer dans l'enseignement supérieur long sélectif. C'est donc au contraire le prestige augmenté des filières nobles qui permet l'augmentation quantitative des filières inférieures.

L'OCDE apprécia le travail et édita le rapport dans un ouvrage qui fut également traduit en anglais (Cibois *et al.* 1976).

### La représentation en surface des tableaux croisés

Du point de vue méthodologique et comme les commanditaires de l'enquête ne souhaitaient pas voir utiliser d'analyse des correspondances mais en rester aux tableaux croisés, je mis au point un système de visualisation des tableaux croisés que j'appelle la représentation en surface. Ce procédé que j'ai exposé dans *L'analyse des données en sociologie* (1984 : 22-43), est une modification de ce que J. Bertin appelle les "matrices pondérées" (Bertin 1967). Cela consistait pour lui à représenter en surface l'effectif de toutes les cases d'un tableau : mon apport à été de représenter seulement ce qui est pertinent dans un tableau c'est à dire ce qui s'écarte de la valeur théorique sous l'hypothèse d'indépendance. De ce fait, on voit immédiatement sur la représentation graphique si l'on a affaire à une attraction (dans ce cas il y a une surface au-dessus de l'indépendance), à une répulsion, un déficit (surface au-dessous de l'indépendance) ou si l'on est dans la situation d'indépendance (pas de surface représentée). La seule information perdue est ce qui correspond à la moyenne d'une des distributions marginales, mais cette information peut être facilement donnée par un graphique en forme de "camembert"<sup>9</sup>.

---

<sup>9</sup> Les anglo-saxons, dont certains trouvent nos fromages trop agressifs, parlent de graphique en part de tarte "pie-chart".

J'ai dans la suite amélioré ce type de graphique d'abord en le formalisant (Cibois 1983b); en l'utilisant pour la représentation en surface de la contribution au Khi-deux d'une case (Cibois 1984), en permutant lignes et colonnes du tableau afin de le "diagonaliser"<sup>10</sup>. Ce concept de diagonalisation étant à prendre dans le sens de J. Bertin (1967) bien que finalement cette opération corresponde au même résultat que la diagonalisation d'une matrice de variance/covariance pour en extraire les valeurs propres. En effet la diagonalisation mathématique a pour résultat de ranger l'information pertinente (les valeurs propres) sur la diagonale d'une matrice et pour J. Bertin, l'opération consiste à mettre sur la diagonale du tableau les forts écarts à l'indépendance, simplement en permutant lignes et colonnes du tableau. J'ai montré dans ma thèse (Cibois 1980 : 153-176) que les deux opérations étaient tout à fait comparables dans leurs résultats et que ce que faisait J. Bertin de manière empirique, l'analyse factorielle le faisait d'une manière formalisée<sup>11</sup>.

En utilisant ce résultat, je commence par faire aujourd'hui une analyse factorielle<sup>12</sup> afin de dégager l'ordre du premier facteur qui va servir à réordonner les lignes et les colonnes du tableau d'origine. Ensuite on peut faire une représentation en surface et les grandes masses d'écarts à l'indépendance se trouveront sur une diagonale du tableau<sup>13</sup>.

Sur cette question mon apport théorique a été de mettre en avant dans un tableau croisé l'écart à l'indépendance dont j'ai découvert ensuite qu'il jouait un rôle clé :

- dans l'interprétation des tableaux croisés car son signe permet de repérer attractions et rejets, - dans l'interprétation des plans factoriels où les correspondances en question se retrouvent sous la forme de conjonction ou d'opposition (et de quadrature pour l'indépendance),

- dans l'interprétation des similitudes entre lignes (ou respectivement entre colonnes) qui ont des profils d'écarts similaires dans un tableau et qui se trouvent proches dans un plan factoriel.

A cette époque, j'avais bien vu le rôle clé de l'écart à l'indépendance, mais j'étais resté dans une perspective "morphologique" : je m'intéressais à la forme des écarts. La prise en compte de l'intensité de l'écart sera un apport ultérieur (Cf. chap.8). Je formalisais ainsi la pratique habituelle des sociologues quand ils utilisent des tableaux croisés et se servent des pourcentages en ligne et les comparent avec le pourcentage en ligne de la marge.

Seule la pratique usuelle de la comparaison entre deux pourcentages en ligne entre eux n'est pas prise en compte par cette formalisation mais cette pratique pourtant courante peut être légèrement inexacte. Soit par exemple le tableau suivant issu d'une enquête fictive sur une population d'un millier de cadres à qui l'on demande s'ils accepteraient ou non une mobilité géographique si les conditions financières étaient satisfaisantes. On croise cette question avec le niveau d'étude des cadres en distinguant ceux qui ont Bac 3 et +, les niveaux Bac, les sans-bac. On a les résultats suivants :

	Acceptation		Total
	Oui	Non	
Bac+3	362	238	600
	60,3	39,7	100

<sup>10</sup> Technique rendue opérationnelle dans le programme REPFAC de Tri-deux.

<sup>11</sup> J.P.Benzécri a d'ailleurs montré quelles en étaient les raisons (1973 : 261-287).

<sup>12</sup> non pondérée pour avoir les grandes masses, cf.chap 8.

<sup>13</sup> On peut en option laisser l'ordre d'origine quand il est pertinent.

Bac	128 64,0	72 36,0	200 100
Non-bac	10 5,0	190 95,0	200 100
-----			
Total	500 50,0	500 50,0	1000 100

Au vu des pourcentages en ligne, l'interprétation standard consiste à dire que l'acceptation est la plus forte pour ceux qui ont le niveau bac et que viennent ensuite ceux qui ont le plus haut niveau. Cependant cette interprétation est fallacieuse du fait des disparités de marge : en effet calculons pour les deux cases, bac et bac+3 l'écart à l'indépendance. On a les résultats suivants :

	Observation	Efectif Théorique	Ecart à l'indépendance
Bac+3	362	300	+62
Bac	128	100	+28

Pour étudier la liaison, plaçons-nous dans le cas où elle serait maximale : du fait des marges, pour la case Bac+3, on ne peut mettre que 500 personnes. Si c'était le cas, l'écart à l'indépendance serait de 200. Par rapport à ce maximum, l'observation n'est que de 62 soit en pourcentage 31%.

Pour la case Bac seul, le maximum serait de 200 soit un écart à l'indépendance de 100 dont l'observé ne représente que 28%.

On voit dans ce cas que la liaison est plus forte avec les Bac+3 qu'avec les bac seul. Il peut donc exister des cas où la lecture classique de la différence de proportion lue comme une différence d'intensité de liaison peut être fallacieuse. Il s'agit d'un phénomène assez marginal : sur les 201 tableaux possibles suivant :

490	110	489	111	362	238	291	309	290	310
0	200	1	199	128	72	199	1	200	0
10	190	10	190	10	190	10	190	10	190

seuls les huit ou la première case va de 367 à 360 présentent le phénomène d'inversion. Sans envisager de probabilisation, on voit cependant que cela ne correspond qu'à environ 5% des cas, ce qui est un risque assez acceptable. Bien souvent il suffit de respecter la précaution empirique souvent évoquée comme quoi il faut tenir pour équivalents des pourcentages qui ne diffèrent que de quelques points.

Ce résultat qui complète la formalisation de la lecture des tableaux croisés a été élaboré à l'occasion du présent rapport.

## La demande d'éducation

Le rapport précédent ayant été accepté, l'OCDE me confia une enquête supplémentaire sur la demande d'éducation et je pus ainsi empiriquement examiner diverses enquêtes sur la question. En me situant dans la perspective de *L'inégalité des chances* (Boudon 1973) et en considérant que la demande d'éducation est une demande de statut social, je montrai qu'il n'y avait pas de déconnexion entre le système scolaire et le système des professions. Un autre aspect intéressant de cette recherche est qu'elle manifestait bien à partir des données disponibles (CEREQ, FQP) que la position sociale est liée au plus haut niveau d'enseignement général obtenu et que l'effet de l'enseignement professionnel qui vient après est d'abord de favoriser l'insertion sur le marché du travail. Les recommandations données en fin de travail consistaient à demander la création d'un enseignement court à vocation professionnelle à la sortie des enseignements longs qui n'en disposent pas comme en fin de Deug ou de licence.

Du point de vue méthodologique, je pus utiliser la grande enquête de Reuchlin et Bacher (1969) sur l'orientation en fin de 3<sup>e</sup> ce qui me permit une bonne confrontation entre des résultats factoriels et la lecture en surface d'un tableau croisé. Cette enquête (Cibois 1976) me permit aussi de d'éprouver la validité de la représentation des évolutions chiffrées par le biais d'une représentation logarithmique.

### 80% de bacheliers

Dans le vol qui me ramenait du Paraguay où j'avais été avec des collègues belges présenter des techniques d'analyse de données pour le compte de la communauté européenne, j'exposais à mon voisin, Jean-Jacques Droesbeke, mes hésitations sur le fait du hasard dans l'évolution de la courbe des bacheliers. Celui-ci fut fort intéressé et me proposa ses services pour étudier avec précision les résultats empiriques : de là naquit une collaboration qui allait conduire à un article sur l'évolution des bacheliers (Cibois *et al.* 1988a). Le but de l'opération fut de faire un ajustement entre la courbe des bacheliers et la loi logistique : cet ajustement permettait de prévoir un maximum aux alentours de la moitié de la classe d'âge.

Comme cet article coïncidait avec le lancement de la campagne ministérielle de 1980 pour l'obtention d'une proportion de 80% de jeunes au niveau du bac en l'an 2000, une confrontation se révéla intéressante. La courbe des bacheliers subit une forte inflexion au début des années 1985. Pour obtenir ce résultat, il suffisait de déverrouiller le passage en seconde ce qui augmentait mécaniquement les effectifs et les reçus en terminale : le nombre d'élèves par classe se mit à croître et les structures eurent du mal à accuser le coup. Les grèves de lycéens et d'étudiants vinrent rapidement sonner l'alarme sur les possibilités d'accueil de ces nouveaux bacheliers.

En 1989 dans un article du *Monde*, je proposais une prise en compte des volontés de promotion d'une plus grande partie de la classe d'âge en pronostiquant que l'évolution antérieure conduisait à la moitié de la classe d'âge, que la volonté ministérielle voulait arriver à 74% de reçus et donc que l'on avait de bonnes chances que la réalité obtenue se situe entre les deux chiffres. Nous en sommes actuellement à 50% mais le chiffre de 80% ne constitue plus un objectif prioritaire étant données les difficultés d'hébergement d'élèves et d'étudiants.

Plusieurs considérations sont à tirer de cette expérience :

- que la notion de niveau, puisqu'elle se réfère à un instrument de précision, inclut une idée de précision et non d'à peu près. Pour être au niveau olympique, il faut avoir franchi une

hauteur bien précise. Par contre dans le jargon administratif de l'éducation nationale, le niveau du bac, c'est la terminale.

- que cet article a été critiqué récemment pour des raisons techniques (Mear et Merle 1991), en particulier sur la mauvaise qualité de la corrélation pour juger d'un ajustement et sur le fait qu'une prévision dans ce domaine ne peut pas prévoir un changement de politique. Ces critiques sont tout à fait pertinentes mais il faudra voir cependant dans quelques années si la politique actuelle, qui ne considère plus les 80% comme une priorité, ne rendra pas les prévisions actuelles également fautives.

## Analyses récentes

Une réexploitation des données du Panel DEP des élèves du Ministère de l'éducation nationale (situation en 1991) pour le compte d'un travail avec F.de Singly (1993), m'a permis de mettre au point une nouvelle technique dite des "profils". En effet la pratique du sociologue est de regarder dans plusieurs tableaux croisés le comportement d'une modalité qui l'intéresse et ainsi d'arriver à en rendre compte, à l'expliquer. Cette pratique conduit en général à ce genre littéraire hautement soporifique qu'est le commentaire de tableaux croisés où une modalité est ainsi comparée de tableaux en tableaux.

Pour formaliser cette pratique, plusieurs décisions doivent être prises :

- définir l'univers de référence : quels sont les autres modalités de l'enquête que l'on souhaite voir être associées avec la modalité dont on veut rendre compte. En étant large, on peut fixer à environ 200 le nombre de modalités soit de comportement, soit de statut, que l'on désire croiser avec une modalité ;

- définir l'indicateur de liaison : prendra-t-on l'écart à l'indépendance comme dans la première version de mon logiciel Tri-deux ? Ou un indicateur qui tienne compte des marges. C'est cette dernière option qui est retenue à travers le PEM, Pourcentage de l'Ecart Maximum ;

- définir la fiabilité de la liaison observée : une liaison peut être due à des effectifs non significatifs. Pour avoir une idée de cette fiabilité, on calcule un Khi-deux (muni de la correction de Yates) et l'on indique les seuils standards de significativité (10%, 5%, 1%).

Pour juger de l'intensité de la liaison, je mis au point un indice, le PEM (Pourcentage de l'Ecart Maximum) qui s'interprète comme son nom l'indique comme le rapport entre l'écart à l'indépendance *observé* et l'écart à l'indépendance *qu'il y aurait si la liaison était maximum*. Cette indice a le mérite d'être immédiatement intelligible par le sociologue : il a le défaut de devoir nécessiter quelques calculs mais comme il est programmé, ce défaut n'en est plus un. Cependant, il faut juger cet indice par rapport à un indice à la même visée appelé "taux de liaison" (Rouanet 1987 : 158) et qui avait été proposé auparavant par J. Bertin sous une forme voisine (centré à 100) sous le nom de "rapport de distorsion" (Bertin 1967 : 225), puis sous une forme identique sous le nom de "khi" (pour montrer son lien avec le khi-deux : Bertin 1977 : 65). Comme cette question n'a pas encore été discutée, un excursus sur ce point s'impose.

## Comparaison PEM taux de liaison

Le PEM est le rapport entre l'écart à l'indépendance observé (dans la suite "écart") à l'écart à l'indépendance qu'il y aurait si la liaison était maximum en tenant compte des marges ("écart- max"). Le taux de liaison de H. Rouanet a le même numérateur mais le dénominateur est l'effectif théorique ("théorique") sous l'hypothèse d'indépendance (et qui tient compte aussi des marges).

Comme ces deux indices ont même numérateur, ils sont nuls quand le numérateur est nul, c'est à dire dans le cas d'indépendance, ils sont négatifs si l'écart est négatif, positifs dans le cas contraire (qui correspond à l'attraction).

Examinons le cas de l'attraction qui est le plus intéressant pour le chercheur : le taux de liaison examine ce que représente l'écart par rapport à la situation d'indépendance. Si l'écart est inférieur au théorique, alors la contribution au khi-deux de la case sera inférieure à l'écart, l'information apportée par la case sera faible tandis que si l'écart est supérieure à l'écart, la contribution au khi-deux sera supérieure à l'écart et l'information apportée sera grande. On peut donc donner une première interprétation du taux de liaison en indicateur d'information apportée par la case : faible si le taux est inférieur à l'unité, forte au-delà.

Si le minimum 0 est bien défini et correspond à l'indépendance, le maximum ne l'est pas : des fortes valeurs correspondent à des cases apportant beaucoup d'information. On doit davantage parler d'indice d'information que d'indice de liaison même si pour un théorique donné, la liaison est d'autant plus forte que l'écart est fort.

Dans le cas du PEM, l'indice est borné à 1 pour la liaison maximale : on a donc réellement affaire à un taux de liaison. On pourrait d'ailleurs le retrouver avec le taux de liaison précédent si on le normait par le taux maximum. En effet :

taux = écart / théorique,  
 taux maximum = écart-max / théorique  
 et taux / taux maximum = écart / écart-max = PEM.

## Exemple d'utilisation

Je voudrais montrer un exemple d'utilisation de cette technique en comparant deux populations scolaires qui présentent un certain nombre de points communs puisqu'il s'agit d'élèves de faible niveau, mais qui n'ont pas la même attitude vis-à-vis de la lecture. En effet, comme on connaissait depuis longtemps la liaison qui existait entre le niveau scolaire et la lecture, il a paru intéressant de "décristalliser" cette liaison (Singly 1993) et de regarder les bons élèves qui, à l'encontre de ce qui est habituellement attendu, lisaient peu, ou inversement les élèves de faible niveau qui avaient une attitude plutôt favorable vis-à-vis de la lecture.

C'est en faisant cette exploration que j'ai vu que parmi les élèves qui étaient tout à fait d'accord pour dire que la lecture était classée comme leur loisir favori, se trouvaient des élèves d'origine étrangère défavorisée (originaires du maghreb, de turquie ou d'afrique). Pour expliquer cette liaison qui semblait surprenante, j'ai entrepris une comparaison systématique entre cette catégorie et une catégorie équivalente de Français, celle des élèves français de niveau scolaire équivalent à celui des étrangers. Du Panel DEP du Ministère de l'éducation qui suit des élèves à partir de la 6e, on tire donc deux populations comparables par leur faible niveau scolaire mais qui les uns sont français d'origine et les autres sont étrangers. A cette fin on croise la nationalité (NAT0=français, NAT1= maghreb + turquie + afrique) avec le niveau scolaire en 4 positions (du niveau le plus faible au plus élevé), on a :

	NIV0	NIV1	NIV2	NIV3	Tot.
NAT0 france	477	448	382	367	1674
	28,4	26,7	22,9	21,9	100
NAT1 maghreb	51	39	12	10	112
	45,6	34,8	10,7	8,9	100

On constate que 80,4% des "étrangers défavorisés" sont en niveau 0 ou 1 : on utilisera donc comme population française de comparaison les 477 + 448 = 925 français de niveau 0 ou 1 (population codée FN02 c'est à dire Français de Niveau faible)

La confection d'un "profil" consiste à croiser une modalité donnée (étrangers défavorisés puis français défavorisés), avec toutes les autres modalités d'un univers de questions qui englobe tous les domaines de l'enquête : les pratiques éducatives, les goûts, la relation à la télévision, à la lecture et aux loisirs, les occupations des jours de congé ainsi que toutes les variables de la vie scolaire concernant le cursus passé, le cursus en cours ainsi que la réussite.

Pour chaque croisement de la modalité dont on fait le profil avec toutes les autres modalités de l'univers de référence, on donne :

- le PEM centré sur 100 (0 liaison négative maximum, 100 indépendance, 200 liaison positive maximum),

- le Khi-deux associé à l'écart à l'indépendance de la case (avec la correction de continuité si un effectif théorique est inférieur à 5 et ramené à zéro si l'écart à l'indépendance est inférieur à 0.5),

- la mention de significativité du test (qui correspond toujours à un tableau 2 x 2 à 1 degré de liberté). Il y a 3 options : "ns" pour non significatif, "\*" pour significatif à .1, "\*\*\*" pour significatif à .05 et "\*\*\*\*" pour significatif à .01.

Profil des étrangers défavorisés :

Nom	PEM	Khi2	Seuil	Intitulé
Q72	126	15.48	***	un retard, c'est grave

Il y a une liaison positive, elle représente 26% de la liaison maximum et elle est significative. Un retard en classe est considéré comme quelque chose de grave. Ceci est un indice d'une "bonne volonté scolaire" confirmée par la liaison avec l'intérêt fort pour le français, bien que le niveau apprécié par l'enseignant soit jugé faible et que le retard scolaire apparaisse déjà :

Q9D1	123	12.72	***	intérêt français beaucoup
NIV0	122	13.86	***	appr.niv. faible
AG64	116	23.02	***	âge entrée 6e : 13 ans
CL23	109	14.79	***	ni 5e ni 6e en 2e année du panel
CL32	121	15.39	***	5e en 3e année du panel

Les parents, ouvriers, non diplômés, ne parlant pas français à la maison ne sont pas à même d'intervenir dans la scolarité de leur enfant :

30 2	113	4.76	**	vérif.parents jamais
CS 6	147	56.79	***	ouvriers
NBE5	150	273.35	***	5 enfants et plus
52P1	124	24.57	***	père aucun diplôme
52M1	113	4.66	**	mère aucun diplôme

A811	146	191.41	***	on ne parle pas français
A732	127	31.33	***	enfant n'a pas une chambre

La télévision joue un grand rôle :

36A1	109	4.15	**	télé le matin avant cours
36C1	121	8.00	***	télé après les cours

mais pas le sport :

50D0	136	8.56	***	mercredi sport rarement ou jamais
------	-----	------	-----	-----------------------------------

mais ce qui est plus surprenant, c'est la vision de la lecture où l'on retrouve le message de l'école, autre signe de bonne volonté scolaire :

38C1	113	9.85	***	aime encyclopédies beaucoup
39D1	122	5.97	**	lire c'est amusant : tout à fait d'accord
39E1	109	3.07	*	lecture loisir favori : tout à fait d'accord

De ce fait la bibliothèque est un lieu fréquenté assidument :

USB2	116	24.39	***	bibliothèque le mercredi : toujours ou souvent
------	-----	-------	-----	--

A travers ce profil on devine le rôle positif en matière d'intégration que peut jouer l'école par le biais de la lecture et de la fréquentation de la bibliothèque, du fait de cette écoute attentive du message donné par l'école, message bien reçu, sans même la pression des parents et alors même que le niveau scolaire en français comme ailleurs est faible.

Il n'en est pas de même pour le profil des français défavorisés qui ont en commun d'avoir un faible niveau mais qui ont des attitudes toutes différentes :

Le niveau est faible :

NIV0	171	410.18	***	appr.niv. faible
CL13	136	39.81	***	redouble 6e en 89

Le milieu est le même que précédemment :

CS 6	110	9.10	***	CSP ouvriers
52P1	110	5.30	**	père aucun diplôme

mais inversement les parents surveillent les horaires et les devoirs :

22 1	111	12.55	***	aide parents toujours ou souvent
30 1	109	4.04	**	vérification des parents tous les jours
17 1	119	7.21	***	parents décident travail
20 2	104	1.54	ns	se couche avant 21h30

mais l'attitude de l'enfant vis-à-vis de l'école est beaucoup moins respectueuse des "valeurs" scolaires. Tant en ce qui concerne l'attitude générale :

Q4 1	112	11.22	***	oubli d'un livre souvent ou parfois
------	-----	-------	-----	-------------------------------------

mais également en ce qui concerne le français et la lecture :

Q9D4	129	12.28	***	intérêt franç. pas du tout
32C0	118	35.46	***	pas inscrit bibliothèque
38B3	128	31.61	***	aime lire livre non
38C3	105	4.11	**	aime encyclopédies non
39A1	126	12.55	***	lire difficile tout à fait d'accord
39B1	129	14.89	***	lit pour étudier tout à fait d'accord
39D4	133	13.93	***	lire amusant pas du tout d'accord
40 2	125	14.40	***	n'a pas lu un livre depuis 3 mois
USB0	118	36.52	***	bibliothèque rarement ou jamais

Il n'y a pas de pratique de la lecture ou plutôt elle est associée à la pratique scolaire qui est surveillée par les parents.

Il y a vraiment une différence entre les deux profils sur la lecture entre les étrangers défavorisés qui "écoutent" le message véhiculé par l'école et y croient tandis que les français défavorisés ne l'entendent pas : seuls leurs parents croient à la réussite scolaire alors que les parents étrangers ne s'en soucient guère. A l'inverse ce sont les enfants étrangers qui considèrent que c'est grave d'arriver en retard alors que les élèves français oublient souvent leurs livres.

Les enfants étrangers croient à ce qu'on leur dit à l'école et passent à l'acte en ce qui concerne les recommandations de lecture alors que les français n'entendent rien de ce qui leur est dit, si ce n'est par le biais de la pression parentale. Pour les enfants français défavorisés, l'école est contre-productive en matière de lecture alors qu'elle est efficace pour les enfants équivalents étrangers.

On voit sur cet exemple que ce qui nécessitait un tableau ne nécessite plus qu'une ligne et qu'on voit tout de suite s'il y a ou non une liaison, dans quel sens, et si cette liaison est significative. On peut ainsi déterminer un profil de chaque modalité qui est en quelque sorte expliquée par les liens qu'elle a avec les autres modalités de l'enquête. Cette explication n'est pas plus circulaire que dans un dictionnaire où un mot est expliqué par d'autres mots du dictionnaire. C'est le réseau sémantique qui est signifiant.

La fécondité de cette technique des profils vient de la facilité de sa mise en oeuvre logicielle : elle permet cette exploration systématique que le sociologue a toujours envie de faire mais qui lui était rendue difficile par la masse des tableaux générés. Désormais, pour toute modalité, il est possible de faire une exploration systématique, sans a priori, qui se révèle rapidement fructueuse ou non.

## Conclusion

Si mon parcours de sociologie de l'éducation s'est révélé fructueux au plan méthodologique car il m'a permis de formaliser la pratique de la lecture des tableaux croisés, de l'automatiser et de l'améliorer ; sur le plan des résultats, je le trouvais assez décevant car toutes ces enquêtes avaient été faites à partir d'interrogations sociales précises et donc limitées. Il ne s'agissait pas de replacer le problème de l'enseignement supérieur court, de la demande d'éducation dans l'ensemble des problèmes de la société, de faire l'inventaire du débat sur la question, mais de répondre d'une façon précise à une question précise. Sur ce dernier point, je pensais être arrivé à des résultats fiables mais qui, s'ils avaient un intérêt social, n'avaient pas automatiquement un intérêt scientifique.

Les instruments méthodologiques mis au point permettent aux chercheurs qui utilisent des enquêtes de répondre rapidement à des questions précises par le biais du tableau croisé simple ou par une association de tableaux croisés sous forme de profils. Cependant, il manque encore une perspective apportée par l'analyse des données, celle qui permet d'avoir une vision d'ensemble sur les répondants d'une enquête, d'en faire une typologie. C'est cette perspective qui va être envisagée maintenant.

## Chapitre 3 : analyse des données

### L'expérience du LISH

Ayant été engagé comme "secrétaire scientifique" lors de la création du Laboratoire d'informatique pour les sciences de l'homme (LISH-CNRS) en 1975, mon rôle fut d'aider les chercheurs fréquentant le service de calcul de la Maison des Sciences de l'Homme à utiliser l'informatique. Je tirai les leçons de mes expériences passées : l'informatique n'est correctement utilisée que s'il y a de la part de ses utilisateurs un investissement tant sur l'informatique elle-même que sur les techniques statistiques utilisées.

Pour ce qui est de l'utilisation de l'informatique, je soutins, en collaboration étroite avec la responsable du service, Monique Renaud, que le centre de calcul ne pouvait correctement fonctionner qu'en "libre-service", c'est à dire avec un mode de fonctionnement où les utilisateurs créaient eux-mêmes leurs cartes perforées et sortaient eux-mêmes les listings qui en résultaient. Cela pouvait sembler peu de chose mais cette organisation supprimait l'ancien système du travail "à façon" où un ingénieur prend (mal) en charge des travaux qui sont ensuite examinés (sans investissement suffisant) par un chercheur.

La nouvelle organisation permettait au chercheur de faire lui-même les travaux informatiques à la condition qu'il y soit formé : des stages nombreux permirent à des utilisateurs de s'initier aux joies du système d'exploitation IBM 360 du Circe et le LISH grâce à cette nouvelle organisation devint rapidement un centre de calcul performant et très fréquenté. Cependant, il ne suffisait pas de savoir se servir matériellement de la machine, il fallait encore maîtriser les techniques statistiques utilisées. Comme bien souvent les utilisateurs venaient avec des enquêtes qu'ils voulaient dépouiller, j'éprouvais le besoin de mettre au point une méthode d'enseignement et des stages de formation adaptés au chercheurs afin de leur faire comprendre l'analyse des correspondances qui était la technique la plus à même de voir l'ensemble des données dont ils disposaient.

Je repris à cette fin le mode d'exposition mis au point par G. Guilbaud et avec un jeu d'exemples facilement compréhensibles par des chercheurs de toutes disciplines, je commençais à faire des stages de quatre jours, à partir de 1979 (et jusqu'en 1989) à raison de 20 à 30 chercheurs par stage quatre à cinq fois dans l'année. En effet en plus des stages trimestriels du LISH, je fus amené à faire ce stage pour l'EHESS, pour la formation permanente du CNRS, pour des écoles d'été du CNRS à Grenoble et à Lille, à l'INRP, à l'ENS Fontenay, à l'INRA (Paris) et pour diverses institutions ainsi qu'à l'étranger.

Mes nombreux contacts avec des chercheurs en situation de recherche me firent voir aussi les limites de l'analyse des correspondances appliquée au dépouillement d'enquête et ce fut pour moi l'occasion de recherches méthodologiques que je proposais dans ma thèse de doctorat de troisième cycle soutenue en juillet 1980 sous la direction de R. Boudon (les autres membres du jury étant MM. M. Barbut et H. Rouanet), sous le titre de "la représentation factorielle des tableaux croisés et des données d'enquête : étude de méthodologie sociologique" (1980).

### Thèse de troisième cycle

L'apport de cette thèse s'est fait dans trois directions :

1) couplage des techniques de J. Bertin avec l'analyse des correspondances. J'ai tout d'abord formalisé la technique de la représentation en surface des écarts à l'indépendance (étude des limites des indices). J'ai montré que la représentation en surface pouvait être utilisée pour la représentation de la contribution au Khi-deux d'une case et qu'il devenait ainsi

possible de représenter un facteur ; que la technique Bertin des matrices pondérées était en quelque sorte une analyse factorielle manuelle et qu'il était tout à fait intéressant de coupler les deux méthodes pour éclairer l'une par l'autre, en particulier, après une analyse factorielle, pour retrouver d'une autre manière dans les données ce qui avait été découvert ;

2) que ce retour aux données était indispensable pour que le sociologue, qui avait peu de chance d'être un spécialiste de l'analyse factorielle, puisse l'utiliser sans risque en prouvant par des techniques simples, ensemblistes, ce qu'il avait trouvé dans un plan factoriel. Je montrais qu'il était facile de construire des "variables idéal-typiques", dont la logique est simplement de compter le nombre d'individus qui possèdent un nombre plus ou moins important de modalités spécifiques d'une partie d'un plan factoriel. Le nom choisi faisait évidemment référence à Weber et à sa notion d'ensemble intellectuellement cohérent même s'il est peu attesté statistiquement.

Dans le même ordre d'idée, c'est là que pour la première fois j'ai envisagé (en utilisant la recherche précédente sur les maximums des indices de représentations en surfaces des écarts), la procédure qui consiste à rechercher quelle pourrait être, pour un tableau à marges données, la situation de dépendance maximum et à comparer par exemple le khi-deux observé avec le khi-deux maximum. En utilisant un algorithme fourni par G. Guilbaud et qui a une solution unique quand on dispose d'un ordre sur les lignes et sur les colonnes, j'ai montré qu'on pouvait ainsi construire un tel indice qui permet de voir le degré de liaison mieux que ne le fait le V de Cramér dont c'était un perfectionnement ;

3) enfin dans une dernière partie j'ai tenté une analyse sociologique du développement de l'analyse des données. J'ai montré que les motivations du créateur de l'analyse des correspondances, J.P. Benzécri étaient sans aucun doute de nature religieuse et que cela expliquait un certain nombre de choses et en particulier pourquoi il insistait tant sur l'examen des facteurs et non pas sur les données elles-mêmes. C'est qu'en effet pour Benzécri, les données sont invalidées par les *a priori* des chercheurs tandis que les facteurs, fondés sur de vastes ensembles, permettent de retrouver l'essence des choses telle qu'elle a été créée par Dieu. Le rôle du chercheur est de découvrir une réalité préexistante, non d'en construire une par des hypothèses hasardeuses.

Comme cette insistance sur les facteurs était prise pour un résultat mathématique et non comme une position philosophique, élèves et utilisateurs de Benzécri transmettaient cette doctrine comme quelque chose d'indiscutable. L'intérêt de mon analyse fut de montrer qu'on pouvait prendre ses distances avec cette pratique et que le retour aux données était une attitude légitime. On pouvait ainsi mettre l'accent sur des techniques diverses comme la représentation d'un tableau en écart à l'indépendance ou la construction de variables idéal-typique qui permettent de vérifier ce qu'on a découvert par une analyse des correspondances.

Je montrais également qu'une partie du succès de l'analyse des correspondances venait de ce que j'appelais l'effet d'"homothétie" qui fait que c'est la *structure* des écarts à l'indépendance et non leur *force* qui est mise en valeur par une représentation factorielle. Ainsi de faibles écarts étaient magnifiés par l'analyse alors qu'ils n'auraient peut-être pas été pris en compte si on les avaient simplement vus dans un tableau croisé. La conclusion n'était pas que les chercheurs travaillaient sur des écarts insignifiants mais qu'au contraire, il n'était pas nécessaire d'avoir des écarts gigantesques pour dire des choses intéressantes.

Je découpais ensuite cette thèse en plusieurs publications qui mirent l'accent sur la présentation de l'analyse factorielle (Cibois 1983a), sur la méthodologie de son utilisation en sociologie (Cibois 1984) et sur les problèmes de l'interprétation de la philosophie de Benzécri (Cibois 1981).

## *L'analyse factorielle*

Le "Que sais-je" paru sous ce titre aux Presses Universitaires de France en 1983 (Cibois 1983a, 3e édition 1991) m'a permis de mettre sous une forme accessible les principes de l'analyse factorielle (composantes principales et correspondances). Le mode d'exposition choisi est celui de G. Guilbaud (décomposition d'une matrice en une somme de matrices de rang un) et la présentation se fait toujours à partir d'exemples numériques sans donner la justification mathématique. On considère que celle-ci est nécessaire pour la validation de la méthode et non pour sa compréhension. Le chercheur doit pouvoir démontrer les mécanismes de la technique statistique qu'il utilise, non les démontrer. Il doit comprendre, du point de vue de l'utilisateur, ce que représente un facteur, une contribution, une valeur propre. Il doit pouvoir comprendre qu'un facteur représente ce qu'ont en commun lignes et colonnes d'une approximation donnée qui entraîne des contributions et des valeurs propres qui sont des part de Khi-deux. De plus le chercheur doit étalonner pour son propre compte la méthode, en plus de la compréhension qu'il doit en avoir. Il doit enfin à partir d'exemples simples puis plus complexes, apprendre à identifier et à interpréter des facteurs et des associations de modalités dans un plan factoriel.

A la fin du livre, on trouvait un retour aux sources de l'analyse factorielle avec la première analyse faite par Spearman en 1904 ainsi qu'une comparaison avec l'analyse factorielle des psychologues en utilisant des données de Reuchlin (Reuchlin 1964).

## *L'analyse des données en sociologie*

Ce deuxième ouvrage, paru en 1984 dans la collection "Le Sociologue" (Cibois 1984, 2ème édition 1990), a été complémentaire du précédent sur plusieurs points. Le "Que sais-je ?" étant déjà sorti, il était inutile de redire ce qui s'y trouvait et c'est pourquoi dans cette nouvelle présentation de la méthode, j'utilisais non plus l'exposition par le calcul mais essentiellement d'une manière graphique. Je voulais utiliser la connaissance que l'on pourrait avoir des techniques graphiques issues de J. Bertin pour assimiler l'analyse factorielle. Je fus là un peu déçu dans la mesure où les gens qui utilisaient les techniques Bertin ne virent pas le pont que j'avais lancé entre leurs techniques et l'analyse factorielle mais, si je n'ai pas eu d'échos officiels, cette présentation purement graphique a peut-être éclairé certains sans que je le sache.

Le lien avec la représentation en surface est développé dans cet ouvrage en partant du principe que s'il existe un ordre dans les lignes et les colonnes d'un tableau, on peut le retrouver dans le premier facteur de l'analyse factorielle de ce tableau. On peut ainsi faire préalablement l'analyse du tableau, trier lignes et colonnes selon l'ordre du premier facteur et représenter en surface les écarts à l'indépendance du tableau permuté : on optimise ainsi la lisibilité des résultats et l'on peut aussi calculer un indice de liaison en comparant la liaison observée avec la liaison maximum. Je nommais dans cet ouvrage un tel indice le PKM Pourcentage du Khi-deux Maximum.

Inversement, je proposais aussi une lecture simplifiée mais rapide des écarts à l'indépendance en observant simplement le signe de l'écart à l'indépendance : cette lecture rapide permet de visualiser la structure globale des écarts à l'indépendance. Il s'agit là de la formalisation d'une pratique courante du sociologue qui souvent fait des approximations indispensables dans un tableau (par exemple en considérant comme identiques des écarts de pourcentage en lignes de quelques points).

Par rapport à la thèse précédente, l'apport de ce livre a été de mettre au clair le concept de méthode "post-factorielle" qui était sous-jacent aux variables idéal-typiques. Le principe en est que l'analyse factorielle permet de se donner des idées lors d'un dépouillement d'enquête

mais que bien souvent, il faut revenir aux données d'origine pour prouver par d'autres méthodes ce que l'on vient de voir. Ceci est nécessité par la raison théorique que l'analyse factorielle ne propose qu'une approximation des données, qui peut n'être que succincte si l'on ne prend que le premier plan factoriel. De plus, alors que l'analyse d'un simple tableau croisé donne un plan factoriel unique, si on change les questions traitées en dépouillement d'enquête, le résultat peut différer assez largement. De plus encore, même si l'on fixe les questions étudiées, du fait de l'utilisation des supplémentaires et des recodages des modalités des questions, les résultats obtenus peuvent sensiblement varier.

Dans l'utilisation de l'analyse factorielle pour le dépouillement d'enquête, toutes ces raisons théoriques et pratiques font que des résultats très divers peuvent être obtenus sur les mêmes données. De ce fait on comprendra que le chercheur, surtout s'il ne dispose pas de cette expérience de la méthode acquise antérieurement et qui lui sert de garde-fou, ait besoin de vérifier ce qu'il vient de voir et qui lui semble assez aléatoire. Il sait bien qu'il a fait de nombreuses analyses et qu'il retient celle qui lui parle, qui le satisfait mais que cette procédure est dangereuse car elle favorise les a priori du chercheur.

Pour le sécuriser, il importe donc qu'après l'analyse factorielle, il utilise des méthodes simples, sans ambiguïté, qui lui permette de dire si ce qu'il a trouvé est un simple artefact ou si cela correspond bien à un effet suffisamment fort pour pouvoir être retrouvé avec des méthodes différentes. Mais il ne faut pas croire que ces méthodes simples, ensemblistes en général, donc à base de tableaux croisés, puissent être utilisées *avant* l'analyse factorielle. En effet si l'on n'a pas de guide pour isoler les tableaux croisés intéressants, on est submergé par l'explosion combinatoire des croisements possibles, et bien souvent on ne croise que les questions que les hypothèses préalables suggèrent. On perd ainsi tout le bénéfice de l'analyse des données qui est de se laisser guider par les données, d'oublier ses hypothèses de recueil pour voir si des configurations inattendues peuvent être suggérées par elles. De ce fait c'est après l'analyse factorielle, et guidée par elle que se font les analyses confirmatoires, d'où le nom de méthode "post-factorielle" où le terme "post" exprime la succession matérielle dans le temps, non le fait que les méthodes employées dépassent l'analyse factorielle.

La première méthode post-factorielle que j'ai proposée a été ce que j'ai appelé la méthode "tri-deux", ce qui signifie "tri au carré" (Cibois 1982a). Elle est post-factorielle en ce sens qu'il s'agit d'un retour sur le tableau de Burt dont on a fait l'analyse factorielle, pour en extraire tous les écarts à l'indépendance positifs : c'est le premier tri. Dans une deuxième étape, tous ces écarts sont rangés en ordre décroissant : c'est le tri sur le tri ou tri au carré. On reporte ensuite sur le graphe factoriel les différentes attractions que signalent les écarts à l'indépendance en commençant par les plus élevés. On tire simplement un trait sur le graphe factoriel pour signaler l'existence d'une attraction à un niveau donné. Comme l'analyse factorielle par définition met en proximité les modalités en attraction, les traits de visualisation se regroupent par secteurs à l'intérieur desquels les attractions sont simultanées. On arrive ainsi à avoir des paquets de graphes non connexes à l'intérieur desquels la connexité est forte dans les meilleurs des cas. Par cette technique, on couple le plan factoriel avec la méthode post-factorielle elle-même ce qui permet au chercheur de mieux comprendre ce qu'il doit voir dans le plan factoriel.

Bien évidemment, une telle méthode n'avait d'intérêt que si elle était facile à mettre en œuvre : elle devait donc être implantée dans un logiciel de dépouillement d'enquête. A cette occasion je rassemblais tous les programmes que j'avais mis au point pour le dépouillement d'enquête en leur donnant comme nom général celui de la méthode qui m'était spécifique : tri-deux. Je présenterai dans le chapitre suivant les problèmes spécifiques, scientifiques et techniques liés à ce logiciel mais je veux montrer dès maintenant les méthodes d'analyse des

données que je fus amenés à créer pour les besoins spécifiques du dépouillement d'enquête. Comme la publication où il est rendu compte de cette recherche est liée à un compte-rendu de colloque non encore publié en français, (en cours de publication directement en anglais chez Sage, Cibois 1993b), je vais la présenter en détail.

## Le tri-factoriel des questions

Le problème est de prendre en compte en même temps les modalités de toutes les questions d'une enquête : comme c'est matériellement impossible, la solution la plus communément employée consiste à répartir les diverses questions de l'enquête en suffisamment de "chapitres" pour pouvoir traiter d'un seul coup toutes les modalités d'un chapitre. Ayant vu le jeu des oppositions entre modalités à l'intérieur d'un domaine, on peut ainsi sélectionner les questions dont les modalités sont les plus contributives aux oppositions. On considère que les autres modalités ne sont pas pertinentes. On fait ce travail pour l'ensemble des chapitres et l'on sélectionne ainsi un nombre plus restreint de questions que l'on sait pertinentes. On peut faire ensuite une analyse globale en n'utilisant que les modalités des questions sélectionnées et avoir ainsi le jeu des oppositions pour l'ensemble de l'enquête.

La difficulté de cette manière de faire réside dans le découpage de l'univers des questions en "chapitres", découpage lié à l'intuition du chercheur, à ses tâtonnements, à ses connaissances préalables du terrain d'enquête. Alors que la détermination du jeu d'opposition de chaque chapitre se fait au vu de l'examen des données, il n'en est pas de même pour la répartition des questions : j'ai donc proposé que le découpage des questions en "domaines" se fasse non pas a priori, mais au vu des résultats d'une analyse.

Pour arriver à ce résultat je propose de travailler d'abord au niveau des questions avant de le faire à celui des *modalités* et de regrouper les questions qui sont liées entre elles. En isolant des groupes de questions liées entre elles, on détermine des domaines et c'est seulement à l'intérieur de ces domaines de questions liées entre elles que l'on fera l'analyse au niveau des modalités. On a donc besoin d'une procédure d'analyse des données pour trouver les questions liées entre elles.

On a d'abord besoin d'un indice de liaison entre questions : en effet si l'on dispose d'un indice numérique qui indique la liaison entre deux questions, il suffira de constituer un tableau qui croise toutes les questions deux à deux et d'en faire l'analyse des correspondances. L'analyse nous indiquera facilement les ensembles de questions qui sont plus liées entre elles : c'est un problème simple d'analyse de données. Le problème réside dans l'indice dans la mesure où dans la majorité des cas, il s'agit de questions nominales et que l'indice ne doit pas être influencé par le nombre de modalités.

La solution que je propose est une extension d'un indice qui se situe dans la logique du  $V$  de Cramer et qui consiste à comparer une situation observée à la situation qu'il y aurait si la liaison était maximum. Par exemple le  $V$  de Cramér met en rapport le Khi-deux observé avec le Khi-deux maximum calculé en fonction des dimensions du tableau. Dans cette logique j'avais proposé le PKM, pourcentage du Khi-deux maximum (Cibois 1984), qui met en rapport le Khi-deux observé avec le Khi-deux qu'il y aurait si la diagonale principale correspondait au plus grand nombre d'individus.

S'il y a des cas où la structure d'ordre s'impose facilement du fait des intitulés, il n'en est pas toujours de même et un ordre sur les lignes et les colonnes est absolument nécessaire pour déterminer d'une façon univoque une diagonale correspondant à la liaison maximum. Mais il est cependant toujours possible de déterminer, au vu des observations, un ordre qui maximise les écarts à l'indépendance positif sur une diagonale simplement par observation de la

structure des écarts et par permutation. En particulier on sait que l'ordre du premier facteur d'une analyse des correspondances répond à ce problème (Benzécri 1973 : 261- 287).

Dans le cas général du croisement de deux questions nominales, il est donc toujours possible de déterminer une diagonale correspondant à des écarts positifs en faisant une analyse préalable du tableau croisé et en permutant lignes et colonnes du tableau selon l'ordre du premier facteur de l'analyse des correspondances. On charge ensuite la diagonale au maximum en fonction des marges et l'on compare le Khi-deux observé avec le Khi-deux correspondant à la liaison maximum.

Un deuxième indice possible qui suit exactement la même logique consiste à s'intéresser aux seuls écarts à l'indépendance positifs et à construire le rapport de la somme des écarts à l'indépendance positifs observés à la somme des mêmes écarts dans le cas de la liaison maximum. Les deux indices ont les mêmes réactions que leurs composantes : l'indice lié au Khi-deux (PKM : pourcentage du Khi-deux maximum) met en relief les écarts correspondant à des marges d'effectifs faibles, l'indice lié à l'écart positif à l'indépendance (PEM : pourcentage de l'écart à l'indépendance maximum) est moins lié aux faibles marges et c'est ce qui nous le fera préférer dans la suite.

Ces deux indices, dans la mesure où ils font référence à un maximum lié au tableau des données, sont indépendant du nombre de modalités, ils varient évidemment de zéro, qui correspond à la situation parfaite d'indépendance à 100, qui correspond à la liaison fonctionnelle.

Si l'on se situe maintenant dans le cas d'un dépouillement d'une enquête importante à 300 questions, il y a 44850 croisements possibles entre ces 300 questions prises deux à deux et il faudra donc calculer 44850 indices pour constituer le tableau où à chaque ligne et à chaque colonne correspond une question et où à l'intersection se situe l'indice de liaison (tableau symétrique et à diagonale non pertinente).

La constitution d'un tel tableau peut sembler irréaliste puisque pour chaque croisement de question il faut :

- calculer les effectifs du tableau croisé,
- en faire l'analyse factorielle,
- permuter lignes et colonnes selon l'ordre du premier facteur,
- maximiser les effectifs de la diagonale qui correspond à la liaison,
- calculer l'indice sur le tableau observé, sur le tableau maximum et en faire le rapport.

En fait, en optimisant les opérations par utilisation d'un algorithme qui n'extrait que le premier facteur, il est possible, en utilisant un microprocesseur 80286 de calculer un indice en un tiers de seconde, c'est à dire de construire un tableau de 300 questions en un peu plus de 4 heures. Le programme qui construit ce tableau (TRIFAC) fonctionne est à la disposition des chercheurs.

Après constitution de ce tableau par tri factoriel des questions, il suffit d'en faire l'analyse des correspondances : les résultats peuvent ensuite être exploités par la méthode Tri-deux en utilisant comme indice entre questions l'indice défini précédemment (PEM).

Devant une enquête importante, il devient ainsi possible d'utiliser plusieurs niveaux d'analyse de données :

1) au niveau le plus général, et après examen des tris à plat et recodages, il devient possible de prendre toutes les questions les plus importantes de l'enquête et de déterminer par

l'utilisation du tri factoriel des questions, des domaines de questions liées entre elles. Eventuellement, à l'intérieur d'un domaine, il sera possible de découvrir des sous-domaines. (Programme TRIFAC et chaîne des programmes Tri-deux)

2) pour un domaine de dimension raisonnable (100 à 200 modalités), on continue à appliquer les techniques classiques d'analyse des correspondances du tableau constitué par tous les croisements entre les modalités. (Chaîne des programmes Tri-deux)

3) au niveau local d'un croisement particulièrement intéressant, il est possible d'utiliser un programme qui visualise en surface les écarts à l'indépendance d'un tableau, éventuellement permuté selon l'ordre d'un premier facteur de l'analyse des correspondances de ce tableau. Ce programme de représentation factorielle (Programme REPFAC) permet une analyse fine d'un seul tableau croisé.

On voit que l'on dispose ainsi des trois niveaux nécessaires à l'analyse des données : le niveau le plus général qui est celui de la *question* et qui permet un découpage sans a priori en domaines disposant d'une certaine autonomie. Le niveau intermédiaire qui permet de découvrir les oppositions entre *modalités*, niveau privilégié de l'étude de la double articulation des variables (variables d'opinion et variables de statut mises soit en actives soit en supplémentaires), et enfin le niveau le plus fin qui nous permet de visualiser des *effectifs d'individus*. Grâce à ces trois niveaux d'analyse on pourra répondre à l'exigence mise en avant plus haut, à savoir de prendre en compte l'ensemble de toutes les données d'une enquête.

Cependant il resterait à prendre en compte plusieurs enquêtes ensemble dans une technique de greffe d'enquêtes qu'il serait intéressante de développer.

\*\*\*

## Chapitre 4 : le logiciel Tri-deux

Dans une démarche scientifique, la création d'un logiciel est un élément d'une stratégie. Si l'on met à la disposition des utilisateurs un ensemble de programmes, c'est évidemment pour leur rendre la recherche plus facile, mais c'est aussi tirer leur usage dans le sens qui semble meilleur par l'utilisation de méthodes que l'on juge plus adaptées. C'est aussi par ce biais que l'on se rend compte si la méthode tient ses promesses : en effet des expérimentations individuelles peuvent ne pas être révélatrices de défauts théoriques mais une diffusion de la méthode auprès de nombreux utilisateurs assure que le banc d'essai sera complet. Proposer une méthode sans le logiciel qui va avec, ce ne peut être qu'une promesse dérisoire, de la poudre aux yeux. Avec un logiciel associé, c'est l'assurance que s'il y a des défauts ou des vices cachés ceux-ci apparaîtront vite et que si la méthode est pertinente et utile, les utilisateurs seront nombreux à vouloir l'utiliser.

Je voudrais dans ce chapitre montrer l'évolution des programmes que j'ai mis au point dans le domaine du dépouillement d'enquête et voir à travers leur évolution comment les aspects pratiques logiciels ont eu des conséquences théoriques. Dans plusieurs cas, je montrerai que des programmes mis au point simplement parce qu'ils semblaient un sous-produit facile à mettre en œuvre, ce sont révélés d'un grand intérêt et on conduit dans la suite à des développements théoriques qu'il était difficile d'imaginer a priori.

## Tri-deux sous IBM 370

La version initiale de Tri-deux tournait sur les ordinateurs du CNRS à Orsay au CIRCE : le système d'exploitation du 370 avec ses nombreuses cartes de contrôle rendait son utilisation difficile pour les utilisateurs occasionnels dont l'apprentissage devait à la fois porter sur des aspects de gestion informatique et sur la technique factorielle elle-même.

Cependant une chaîne de programmes permettait le codage logique de données supposées existantes (modifiés éventuellement sous les puissants systèmes OSIRIS puis SAS) ; le programme créé par un des premiers collaborateurs de Benzécri (E. Yagolnitzer) permettait de faire l'analyse factorielle (correspondances ou composantes principales) du tableau en codage logique ; un programme permettait de créer le tableau de Burt associé où les écarts à l'indépendance positifs étaient relevés puis triés ; et enfin un programme utilisant des possibilités graphiques permettait de reporter sur un plan factoriel les écarts relevés en commençant par les plus importants.

En travaillant sur l'édition de la liste des écarts à l'indépendance, je me suis rendu compte qu'il existait pour chaque ligne du tableau de Burt, une liste facile à éditer, celle des écarts associés à une seule modalité de l'enquête. En triant cette sous-liste et en l'éditant, je crus avoir un simple sous-produit de l'exploration du tableau de Burt. En fait, il me semble aujourd'hui qu'en éditant ce que j'ai appelé à ce moment-là le "profil" d'une modalité, c'est à dire l'ensemble des autres modalités spécialement en lien avec une modalité donnée, j'avais mis au point un outil d'exploration qui modifie profondément la pratique du dépouillement d'enquête.

En effet, avant l'utilisation de l'analyse factorielle, on en était réduit à croiser systématiquement quelques questions que l'on jugeait essentielles comme le sexe, l'âge ou la CSP avec l'ensemble des variables "à expliquer", ce qui générait des masses de tableaux difficilement exploitables. Avec l'analyse factorielle on disposait d'une carte pour se diriger dans ce territoire gigantesque mais cette carte était quelquefois aléatoire et en tout cas imprécise. Elle permettait une analyse globale à peu près correcte mais était entachée d'imprécision et d'incertitude aussitôt que l'on revenait à une analyse locale. Avec les profils, on retrouvait la certitude de résultats issus de tris croisés, on faisait de l'analyse locale puisqu'on recherchait le profil d'une modalité particulière et l'on conservait la vision d'ensemble puisque la sélection des écarts à l'indépendance était faite sur l'ensemble d'un tableau de Burt suffisamment général.

Ces profils étaient rudimentaires puisqu'ils consistaient simplement en la production, par ordre décroissant, des valeurs de l'écart ordinaire à l'indépendance pour le croisement de deux modalités (observé moins théorique). Cependant, n'ayant pas encore correctement évalué l'importance de ces profils, je ne les mettais guère en avant et les présentais comme un simple sous-produit. Je mettais l'accent sur le graphe Tri-deux où ces profils se retrouvent comme étant le graphe associé à une modalité donnée, constitué de toutes les arêtes partant de la modalité en question.

Pendant les stages que j'organisais, j'initiais de nombreux chercheurs à l'usage de Tri-deux mais seul un petit nombre persévérait car il fallait une bonne dose d'obstination pour être efficace avec un système informatique si lourd.

## Tri-deux en micro-informatique : première version

Comprenant bien les difficultés des utilisateurs face à la lourdeur de la grosse informatique, j'accueillis la micro-informatique comme une délivrance. Après avoir participé à une commission du Plan "Informatique pour tous" de 1985 pour les sciences humaines, je

fus nommé responsable d'un des deux "centres de formation des formateurs" spécialisé dans les sciences de l'homme qui, deux années de suite pendant trois semaines forma une vingtaine d'universitaires à l'utilisation de la micro-informatique dans les sciences humaines. A cette fin le LISH fut doté d'une dizaine de Logabax à citoyenneté française théorique mais toute relative mais qui, sans disque dur et avec deux disquettes ont constitué pendant plusieurs années l'outil informatique de base du chercheur en sciences humaines.

Poussé par les besoins du stage, je mis au point une suite de programmes de dépouillement d'enquête. A cette fin, je repris les programmes en Fortran que j'avais développé au CIRCE et devant chaque programme, je rajoutai une étape en conversationnel où l'utilisateur donnait les indications nécessaires au déroulement du programme. Pour l'analyse factorielle, je renonçai au programme de Yagolnitzer dont la complexité était rebelle à la micro, et j'adaptai, avec l'accord de son gestionnaire (J.P. Fénelon) un programme de l'ADDAD appelé ANCORR (Analyse des correspondances).

Pour adapter le programme TRACES qui utilisait le mode graphique pour superposer à un plan factoriel les arêtes d'un graphe représentant les écarts à l'indépendance, je découvris la programmation en Basic puisque le Fortran ne permettait pas le graphique. Ce langage me rappela le PL/1 d'une époque antérieure car il permettait à la fois le calcul et le traitement des chaînes de caractère tout en ayant des possibilités graphiques. Son seul défaut de l'époque était qu'il n'était pas possible de générer un code exécutable dans sa version GWBASIC. Cependant peu après le Turbo-basic de Borland fit du Basic un langage à part entière de la micro et je commençai à l'utiliser plus largement.

Cette première version de Tri-deux avec ses corrections successives, jusqu'à la version 1.4 d'avril 1988 qui ne se modifia plus jusqu'à la version 2 actuelle, fut diffusée auprès d'un millier d'utilisateurs, chercheurs, étudiants et même auprès d'un petit nombre d'utilisateurs qui souhaitaient l'utiliser à des fins professionnelles ou sociales comme le chef de gare d'une importante ville du centre de la France qui voulait explorer les fichiers dont il disposait ou un club de vélo qui voulait se faire une idée de ses adhérents.

Ce chiffre d'un millier concerne les utilisateurs répertoriés nominativement mais beaucoup d'utilisateurs sont en dehors de ce fichier alors que certains qui y sont ne l'utilisent pas. En effet il est tout à fait probable que beaucoup de demandes du logiciel n'ont pas été suivies d'une utilisation dans la mesure où celui-ci, afin de favoriser la diffusion de la connaissance scientifique et technique, était diffusé gratuitement. Mais d'autre part, comme sa reproduction était libre, il s'est diffusé par capillarité sans que j'aie pu m'en rendre compte si ce n'est par des appels au secours téléphoniques.

## Environnement d'assistance et de formation

Un tel logiciel n'a de sens que si les utilisateurs sont aidés dans leur travail : ils le furent d'abord par les stages de formation réguliers déjà cités qui évidemment utilisèrent la micro-informatique pour les travaux pratiques. Mais ils le furent aussi par une assistance technique journalière par téléphone rendue indispensable par la délocalisation des travaux due à la micro. Si cette assistance put sembler lourde à certaines époques puisqu'elle durait une heure par jour en principe et quelquefois plus, en fait je me suis vite aperçu que l'utilisateur de micro-informatique devient beaucoup plus vite autonome que l'utilisateur de grosse informatique. Devoir maîtriser une machine limitée, dont on possède la documentation, sur laquelle on peut faire des essais, qui est chez soi ou dans son laboratoire : tout cet ensemble de conditions fait que le chercheur acquiert vite son autonomie et que l'assistance technique porte souvent plus sur l'analyse factorielle elle-même que sur l'utilisation de la machine.

Je faisais en même temps au LISH une assistance technique plus scientifique puisque pratiquement tous les jours des années 1985 à 1989 je recevais en moyenne un chercheur par jour (ou étudiant avancé) pendant un rendez-vous d'une heure en général. Là, l'assistance était plus scientifique et consistait en général en conseil sur l'utilisation des données : comment coder des données, comment adapter l'utilisation du logiciel pour lui permettre de répondre aux hypothèses de travail et surtout comment interpréter les résultats. Enfin il fallait souvent conforter les chercheurs sur leurs capacités d'interprétation et leur montrer qu'ils pouvaient se rassurer en pratiquant des techniques "post-factorielles" et faciliter ainsi la communication de leurs résultats. Si ce qu'ils avaient vu dans l'analyse factorielle ne leur semblait pas prouvé de façon indiscutable, il était facile de retrouver le même résultat par des techniques simples comme les variables idéal-typiques ou même par de simples tableaux croisés. Je fus amené ainsi à voir de nombreuses recherches, surtout dans le domaine de la sociologie mais aussi en histoire ou en psycho-sociologie.

Ce fut à cette époque qu'un numéro de la revue *Informatique et Sciences Humaines* fut consacré à la Méthode Tri-deux : j'en rappelai les principes, donnai un exemple d'application et un certain nombre de collègues qui l'employait voulurent bien proposer des exemples d'utilisation. H. Rouanet y proposa une note sur la liaison entre les modalités d'un tableau : en effet j'avais à sa demande proposé une première version (observé/théorique) de son taux de liaison (écart/théorique) comme option dans Tri-deux.

## Avatars

Beaucoup d'utilisateurs m'ont demandé pourquoi je diffusais gratuitement ce logiciel : je dois dire qu'au début la question ne s'est guère posée puisqu'il me semblait que la micro-informatique n'était que la continuation de mon travail professionnel de conseil et de formation au LISH. J'expliquais ensuite que l'"intérêt" d'un chercheur n'est pas tellement financier que scientifique et que de ce point de vue-là, c'est en favorisant la diffusion de mon programme que je diffusais par la même les méthodes que j'avais mis au point et donc que j'élargissais ma surface sociale. La demande faite était peut-être aussi une manière de me renvoyer à un rôle professionnel et social d'ingénieur qui était mon titre au CNRS et de m'empêcher d'accéder à celui de chercheur que je souhaitais me voir reconnaître. Il y avait là un problème plus général d'image sociale dont la solution fut de quitter le LISH et de devenir maître de conférences à l'Université Paris V : j'entrai alors de plein pied dans le rôle d'enseignant-chercheur où le problème de la reconnaissance est entièrement lié à la production scientifique et non plus au statut comme auparavant.

Si j'ai quitté le LISH c'est aussi pour une raison liée à l'évolution des centres de calcul en général du fait du développement de la micro-informatique. En effet si dans les années 1975, la seule possibilité pour utiliser l'informatique était de venir dans un centre de calcul comme le LISH, à partir des années 1985 les centres de calcul furent soumis à la concurrence de la micro-informatique. Les chercheurs découvrirent avec ravissement une machine plus facile à utiliser, dont on était le maître et qui permettait de faire pratiquement tout ce que faisait la grosse informatique dans le domaine du dépouillement d'enquête.

Cette évidence ne fut pas immédiatement admise dans les centres de calcul : moi-même je me souviens de mes réticences quand P.O. Flavigny, toujours le premier à étudier de nouvelles solutions techniques, présenta au LISH aux alentours de 1980 le premier micro-ordinateur monté par ses soins et qui ne disposait comme organe de visualisation des résultats que d'un simple compteur numérique à diode. A la question que je posai pour savoir à quoi dans le futur pourrait servir un tel bricolage, il me fut répondu qu'il pourrait faire tout ce qu'on faisait au CIRCE. Je fus tout à fait sceptique et pourtant c'était l'exacte vérité. Si je compare aujourd'hui une abaque prédisant le temps de calcul nécessaire pour une analyse factorielle en

fonction du nombre de modalités actives que j'avais publié dans ma thèse et qui correspondait au CIRCE de l'époque, je m'aperçois que les temps sont analogues, dix ans après, à ceux nécessaires sur un micro équipé d'un processeur 386.

Le même P.O. Flavigny quelques années plus tard forma les chercheurs du CNRS à la micro en leur faisant monter eux-mêmes des ZX81 (à 1K de mémoire), petits appareils géniaux qui m'apprirent à manier le fer à souder et à perdre tout complexe devant la quincaillerie informatique.

Du fait de la micro, tous les centres de calcul (scientifiques ou non) durent opérer une reconversion car les anciens utilisateurs devenaient autonomes grâce à l'investissement à la fois en temps et en argent qu'ils faisaient. De prestataire de service, le centre de calcul passait à celui de conseiller, d'expert, d'assistant. Si individuellement ce passage pour moi correspondait tout à fait à mes vœux et à ma pratique, le LISH ne sut pas faire le passage et usa à la tâche trois directeurs pour effectuer la transition qui est encore en cours.

Si je trouvais normal de distribuer gratuitement un logiciel qui était le fruit de mon travail effectué dans un service public, je fus assez sensible à l'argument que la diffusion que je pouvais en faire était rudimentaire par rapport à celle qu'un éditeur informatique pouvait proposer. Pour cette raison, je répondis positivement aux avances qu'un éditeur qui travaillait à la fois dans le livre et dans l'informatique me fit. Mais l'expérience fut catastrophique : comme mon logiciel avait été écrit par un amateur, il semblait normal à l'éditeur de le faire réécrire par un professionnel dans le langage professionnel d'aujourd'hui, c'est à dire en langage C.

Je retrouvais là le réflexe que j'avais déjà découvert dans mes propres apprentissages informatiques. En effet quand j'eus appris le Fortran, puis le PL/1, on me fit comprendre au LISH que si je voulais être sérieux, il fallait que j'apprenne l'Assembleur, c'est à dire le langage de la machine elle-même et non ces langages de haut niveau qui doivent eux-mêmes être traduits en langage machine. Pour être performant, pour être professionnel, il fallait utiliser l'Assembleur. La même attitude se reproduit avec la micro car le langage C est effectivement beaucoup plus proche du langage machine qu'un langage comme le Pascal considéré cependant comme un langage sérieux. Son seul inconvénient est qu'il faut apprendre à faire ce qui est déjà prévu par un langage évolué comme d'afficher un message à l'écran ou de lire un enregistrement sur un support périphérique. Il me semble que c'est une perte de temps. A cela les professionnels répondent que le C est indépendant de la machine ce qui est exact mais insuffisant car les langages développés comme le Pascal ou d'autres se trouvent aussi maintenant sur la plupart des machines.

En fait j'interprète cette prééminence du C par la liberté d'action qu'il donne au spécialiste qui n'est plus limité par les contraintes du langage évolué. Cette liberté, jugée indispensable par le professionnel, n'est absolument pas nécessaire à celui qui comme moi développe des programmes et pour lequel les langages évolués offrent des possibilités suffisantes et une économie de temps gigantesque. Cependant comme le C est devenu le langage professionnel, beaucoup de chercheurs pensent qu'ils en ont besoin alors qu'il veulent simplement se faire reconnaître comme des professionnels de la micro-informatique. Je pense qu'ils perdent leur temps et qu'ils se font piéger à vouloir acquérir une liberté de programmation qu'ils n'obtiendront jamais car elle suppose une grande pratique et qui leur est complètement inutile pour leurs besoins qui sont satisfaits par les langages évolués.

Ainsi donc Tri-deux devait être réécrit en C et une équipe universitaire informatique fut mise sur le problème : le chargé d'étude de l'éditeur devait assurer l'interface mais je trouvais normal d'aller voir en province ceux qui en étaient chargés. Cela fut considéré comme un

casus belli par l'éditeur qui me rappela les termes du contrat. Je n'avais pas de temps à perdre à discuter et n'insistait pas. Le résultat fut comme prévu mauvais : l'équipe s'est désagrégée sous la pression inquisitoriale du chargé de mission. Le résultat obtenu fut une reconstitution à l'identique de Tri-deux sans aucune des améliorations théoriques qui me semblaient indispensables et avec simplement en plus quelques fonctions d'édition qui venaient d'ailleurs directement d'un autre logiciel. J'attendis une relance de l'éditeur et deux ans après elle n'est toujours pas venue : je pense que le projet a avorté et n'ai aucune envie de faire de nouvelles tentatives dans ce sens.

## Tri-deux deuxième version

Les améliorations à apporter à la première version de Tri-deux sur micro me paraissaient indispensables et elles étaient de deux ordres : ergonomiques d'une part et théoriques d'autre part. Du point de vue de l'ergonomie du logiciel, je me rendis compte que j'avais trop utilisé le conversationnel dans la première version. C'était inévitable dans la mesure où cela avait été une libération par rapport aux cartes perforées mais cela devenait exagéré quand il s'agissait de rentrer sans erreur les caractéristiques précises de 50 questions différentes d'une enquête. Il valait mieux alors créer ces caractéristiques complexes sur un fichier préalable qui serait ensuite lu par le logiciel.

Je fis donc un certain nombre d'améliorations dans ce sens sans investir trop de temps dans des menus déroulants ou autres gadgets à la mode. L'intérêt d'un logiciel scientifique réside dans ce qu'il propose et qu'on ne trouve pas ailleurs. Il n'est pas forcé de ce fait de se différencier par les belles couleurs qu'il utilise où par son utilisation intensive des dernières nouveautés. Il peut se permettre d'être légèrement en retard sur ces points : il doit être au diapason de l'utilisateur ordinaire, non à la pointe de la dernière avancée technique.

J'améliorai Tri-deux en en faisant un logiciel qui se suffit à lui-même en matière de traitement en ce qui concerne les modifications à apporter au fichier de données. Je développai une "boîte à outils" qui permet les recodages, les filtres et des créations de variables nouvelles. Cependant, sans créer un éditeur d'entrée, car il en existe de très performants sur le marché, je développai un interface avec le système le plus utilisé pour entrer des données c'est à dire dBase.

Pour faciliter la lecture des résultats, je rendis possible l'introduction de libellé en clair des modalités qui peuvent par exemple être utilisés dans les plans factoriels. Ces plans factoriels sont maintenant demandés en dehors du programme qui calcule les facteurs. Ce dernier calcul est fait en utilisant un algorithme connu depuis fort longtemps (algorithme des puissances itérées), mais qui n'était pas utilisé car jugé trop lent. Il l'est peut-être, mais comme il extrait les facteurs en commençant par les plus importants, il permet de ne calculer que les 6 ou 7 premiers en négligeant la centaine d'autres qui ne seront pas utilisés. Au bout du compte, on y gagne du temps.

Pour arriver à cette nouvelle version, j'abandonnais le Fortran qui intéresse peu les producteurs de langages et j'utilisais les avancées techniques réalisées par Microsoft sur son Basic. Dire que l'on programme en Basic est plutôt mal vu dans la communauté des professionnels mais en utilisant ce langage on s'approprie une masse gigantesque de travaux développés par de vrais professionnels qui ont perfectionné ce langage et l'ont rendu apte à tous les besoins d'un développeur, y compris pour son intégration dans Windows ou pour l'utilisation des mémoires étendues.

Les développeurs en C y voient là une trace indélébile d'amateurisme que je ne renie pas : je considère Tri-deux dans sa nouvelle version comme le travail d'un amateur éclairé dont l'ambition est de mettre à la disposition des chercheurs des méthodes qu'ils ne pourront pas

trouver ailleurs. Cela suppose de leur part l'acceptation de devoir travailler sans souris ni menu déroulant : l'expérience montre qu'ils le font sans problèmes. D'ailleurs le véritable apport de cette version deux de Tri-deux réside dans ses aspects théoriques et l'on y trouve programmé des techniques originales :

- module REPFAC (REPrésentation FACTorielle des écarts à l'indépendance)

Non seulement il visualise en surface les écarts à l'indépendance d'un tableau mais il peut trier lignes et colonnes selon un ordre issu du premier facteur d'une analyse factorielle préalable. Il y a ainsi une "diagonalisation" automatique des écarts à l'indépendance qui permet une lisibilité plus grande des résultats. Cette technique est directement issue des travaux de ma thèse.

- module ANECAR (ANalyse des ECARts)

L'extraction des facteurs se fait non seulement en utilisant l'analyse des correspondances mais aussi en utilisant les composantes principales non pas sur le tableau brut mais sur le tableau centré mais non réduit des écarts à l'indépendance. Le tableau pouvant être soit le tableau de Burt ou un simple tableau de contingence.

Cette dernière méthode que j'appelle analyse non pondérée des écarts à l'indépendance à été développée pour des raisons que je précise plus loin au chapitre 8. Alors que l'analyse des correspondances met en relief dans ses premiers facteurs les associations de modalités spécifiques à de petits groupes (ce que j'appelle plus loin l'effet de distinction), l'analyse non pondérée permet de mettre en valeur dans les premiers facteurs les liaisons peut-être moins fortes mais liées à des masses plus grandes d'individus.

- module PROFIL

Il permet, pour une modalité donnée, de repérer le profil de toutes les associations de cette modalité avec les autres modalités prises en compte dans un tableau de Burt. La force de cette attraction est dans cette nouvelle version calculée en utilisant le PEM déjà évoqué (Pourcentage de l'Ecart Maximum) et le degré de significativité est donné en faisant un test du Khi- deux sur le tableau considéré.

- module OUTILS

Dans cette boîte à outils, on trouve non seulement des instruments de gestion nécessaires pour des recodages ou des filtres, mais on y trouve directement accessible la possibilité de créer des variables Idéal-typiques utilisées pour des analyses post-factorielles. Il s'agit là d'un outil qu'on ne trouve pas directement utilisables dans les autres logiciels où ce travail nécessite toujours un travail de programmation.

En conclusion on voit que dès à présent, en utilisant la version 2.0 de Tri-deux, un chercheur peut, par importation d'un fichier dBase ou créé en ASCII par ailleurs, faire l'analyse globale de sa population en utilisant diverses formes d'analyse factorielles. Il peut ensuite revenir à une analyse locale soit par des tableaux croisés présentés automatiquement avec l'ordre optimum sur les lignes et les colonnes et d'une manière graphique, soit par des recherches sur des profils de modalités.

Les développements prévus pour la version 2.1 à venir de Tri-deux seront les suivants :

- intégration des graphes Tri-deux : ces types de graphiques sont une confirmation post-factorielle certes, mais d'une analyse globale et c'est pourquoi j'ai tenu en priorité à développer l'analyse locale qui me semble aujourd'hui prioritaire pour le chercheur. En effet le graphe Tri-deux qui ajoute simplement des arêtes sur un plan factoriel, s'il rassure le chercheur ne lui apporte que peu d'éléments nouveaux par rapport au plan factoriel. Dans la

mesure où les chercheurs sont de plus en plus formés à la compréhension du graphe factoriel lui-même, cette confirmation par une méthode post-factorielle devient de moins en moins nécessaire.

- intégration dans le logiciel de tous les programmes séparés développés à sa périphérie : le programme TRIFAC de tri factoriel des questions présenté plus haut et les programmes liés à l'analyse des mots et des textes que je présenterai au chapitre suivant.

- en ce qui concerne les nouvelles méthodes, un module permettant la recherche de styles de comportement ou styles de vie est dès à présent réalisé. Il ne reste plus qu'à l'intégrer dans la nouvelle version. Sur les aspects théoriques posés par cette recherche, cf. le chapitre 8.

\*\*\*

## Chapitre 5 : analyse de contenu

### Sources anciennes : le royaume de Méroë

Mes premiers contacts avec l'étude des textes en utilisant l'informatique remontent à mes premiers travaux pour le compte de l'Ecole Pratique des Hautes Etudes des années 1972 et suivantes où j'étais chargé de continuer l'informatisation des textes écrits dans une langue non encore déchiffrée : le méroïtique. En effet J. Leclant, égyptologue, avait poussé à l'enregistrement des quelques centaines d'inscriptions connues du royaume de Méroë située en Haute-Egypte. Ces textes étaient orthographiés en utilisant une écriture inspiré du démotique égyptien, ce qui rendait la transcription facile, mais le sens des textes restait obscur, on voyait simplement qu'il s'agissait d'inscriptions votives assez stéréotypés qui devait comprendre des invocations aux dieux et des bénédictions.

La confection de concordances et d'index me permit de m'initier au traitement des chaînes de caractères, aux problèmes du tri sur de vastes fichiers et plus profondément au problème du repérage d'éléments connus (ici des formes graphiques) à l'intérieur d'une masse plus large, problème typique de reconnaissance des formes (Cibois *et al.* 1972). Comme je connaissais l'hébreu biblique, langue chamito-sémitique à laquelle le méroïtique semble apparenté, je vis quelques rapprochements intéressants qui séduisirent certains chercheurs mais en horrifièrent d'autres ce qui me convainquit de laisser le dossier à un jeune chercheur dont l'égyptologie était la spécialité et que je formais à l'usage des programmes que j'avais utilisé.

### Analyse de contenu d'un débat : Belle-Ile, été 1984

Après plusieurs années d'assistance méthodologique au LISH, j'avais été frappé par l'importance de la demande de traitement dans le domaine des textes, en particulier du traitement des interviews semi-directifs qui sont souvent la richesse d'un chercheur mais souvent aussi son angoisse. Il a passé beaucoup de temps à interroger et à transcrire et il ne sait plus quoi faire de cette matière première.

Dans ce domaine, je voulais prendre position par rapport à l'analyse de contenu telle qu'on la pratique en sociologie depuis les travaux de Laswell et Berelson (Bibliographie in

Daval 1967) c'est à dire en essayant de compter des thèmes que l'on a déterminé à l'avance. Je voulais critiquer cette manière de faire en m'inspirant tout d'abord des principes de l'analyse des données : il ne fallait pas rechercher des thèmes connus à l'avance mais laisser les thèmes sortir du matériel examiné lui-même. Or en analyse de contenu traditionnelle, les thèmes sont choisis à l'avance par le chercheur : il a certainement de bonnes raisons de le faire mais le texte n'est pas exploité pour lui-même, il sert simplement de support à une investigation qui a ses raisons ailleurs.

Ce désir d'examiner le matériel linguistique dans sa totalité, avait également une autre raison : il me semblait que c'était l'opposition complète des thèmes qui avait valeur explicative, non la présence de l'un ou de l'autre. En reprenant là les principes qui ont fait leur preuve en linguistique, je pensais que c'est le système des oppositions qu'il faut étudier et que pour voir ce système il faut regarder l'ensemble.

De plus, dans les représentations de la société, ce système d'opposition est perçu par les acteurs sous le terme de "débat social". Certes ce n'est pas un concept accessible à l'ensemble des acteurs sociaux, mais il est conscient dans les franges intellectuelles. Par exemple l'opposition gauche/droite en est un exemple tout à fait clair, mais ce n'est pas le seul. On sait que sur beaucoup de problèmes de société, il y a des débats et que, comme le disent les acteurs eux-mêmes, ils "dépassent l'opposition gauche/droite". Ce qui signifie que ces débats sont perçus comme des systèmes d'oppositions où chaque thème est lié à un groupe identifiable, comme dans le débat politique mais ne le recouvrant pas.

Mon choix d'étudier un débat social était donc lié à la prise de conscience que les principes de l'analyse des données portaient en eux une exigence de méthode qui n'avait pas été perçue comme telle. Prendre en compte l'ensemble des données, ce n'est pas simplement une curiosité prudente, un empirisme sans arrière-pensée, c'est bien plus fondamentalement rechercher les oppositions qui *font système* dans un ensemble de données, c'est redécouvrir d'une manière pratique le paradigme méthodologique qui a permis à l'antique grammaire, à la plus récente philologie de devenir la première science de l'homme grâce à la phonologie issue du Cercle linguistique de Prague (Troubetzkoy, Jakobson). Rechercher les oppositions c'est envisager l'utilisation du structuralisme dans des domaines autres que strictement linguistiques.

Cette ambition un peu forte était donc sous-jacente à l'étude d'un matériel que je rencontrai pendant l'été 1984, le compte-rendu par l'ensemble des quotidiens et quelques hebdomadaires d'opinions des événements survenus en juillet à Belle-Ile où un certain nombre de jeunes immigrés, qui y faisaient un séjour dans le cadre de ce qui s'appelait à l'époque l'Opération "anti-été chaud", furent pris à partie par des habitants à la sortie d'un bal. Poursuivis jusqu'à leur camp, ils furent l'objet de menaces et d'injures racistes et durent être activement protégés par les forces de l'ordre. Comme l'affaire touchait à l'insécurité et au racisme, les médias nationaux envahirent l'île, ce qui entraîna une nouvelle réaction des bellilois.

Si j'ai choisi des textes de presse pour mettre au point ma méthode (Cibois 1985), c'est que le débat social s'y déploie puisque chaque journal essaye, pour le compte du groupe social où se recrutent ses lecteurs, d'intégrer dans sa vision du monde les événements qu'il rapporte. De plus, ces tentatives d'intégrations se font sous le mode de l'explication rationnelle, de cette mise en perspective (ou de cette mise en scène) qui permet de comprendre l'histoire qui se déroule sous les yeux des lecteurs.

## La méthode synoptique

D'une manière plus concrète, le corpus de textes de presse recueilli contient 22 textes différents plus ou moins importants et a une longueur totale d'environ 16.000 mots. Le problème méthodologique était donc de déterminer quels étaient les thèmes qui devaient être retenus dans l'analyse de contenu. Si l'on se pose la question directement le problème est insoluble car si un journaliste décrit une situation avec beaucoup de détails, c'est qu'il les juge importants pour ses lecteurs et l'on ne voit pas au nom de quoi le chercheur dirait qu'il n'en est rien et imposerait ses propres catégories.

Cependant, on peut trouver une solution dans l'exigence méthodologique issue du principe de vouloir rechercher le système d'opposition du débat social : en effet pour qu'il y ait trace du débat dans la presse, il faut que plusieurs journaux traitent de façon diverse de la même chose. Inversement, entre deux journaux proches, on pourra dire qu'il y a consensus sur un point s'ils traitent un même thème d'une façon approchée.

Pour repérer les convergences d'utilisation ou les oppositions, je repris la méthode antique de la Synopse, procédé matériel qui permit, par une confrontation physique de versets évangéliques, de se rendre compte que trois auteurs, Marc, Matthieu et Luc avaient beaucoup de points communs et étaient dits pour cette raison "synoptiques", tandis que Jean, disposant de sources autres, peut-être personnelles, ne l'était pas.

Familiarisé par mes études antérieures avec l'exégèse biblique, cette méthode s'imposa à moi avec évidence : il restait à la mettre en œuvre dans le cas présent ce que je fis en employant une méthode très concrète et peu informatique à base de photocopie des textes, de découpage en petits extraits, de collage sur des feuilles séparées. Après quelques tâtonnements qui durèrent d'ailleurs assez longtemps, l'ensemble des textes se répartit en une centaine de thèmes qui non seulement recouvrait les aspects évènementiels, mais aussi les explications et les jugements.

L'utilisation du matériel papier, colle, ciseau, photocopie était liée à la situation de l'époque. Aujourd'hui, grâce aux lecteurs optiques de textes imprimés, il est évident que le logiciel de traitement de texte est un instrument facile d'accès pour réaliser de telles synopses.

On m'a critiqué alors sur la non-formalisation des synopses ainsi obtenues : c'est moi, qui avec ma sensibilité propre déclarait qu'il y avait consensus sur un thème et opposition sur un autre. A cela je réponds qu'il est tout à fait raisonnable de penser que j'ai pu commettre des erreurs ponctuelles d'interprétation mais qu'il est rigoureusement impossible que j'ai pu me tromper partout. Or c'est de l'ensemble des données et non d'une analyse locale que sort l'interprétation générale par le biais d'analyses de données faites sur la totalité du matériel.

## Codage

Une fois la centaine de synopse réalisée, il restait à les coder : le réflexe spontané de l'analyste de contenu eut été de sortir son double-décimètre et de compter la longueur des phrases émises et supposant l'équivalence entre l'importance de la surface émise et l'importance accordée au thème. Je ne repris pas cette manière de faire car une nouvelle fois, je me suis laissé guider par mon but : je ne cherchais pas à mesurer l'importance des thèmes mais le jeu des oppositions. Je voulais voir si à un thème s'en opposait un autre. Concrètement, je devais être capable de repérer la présence mais aussi l'absence d'un thème.

Ce repérage de l'absence est tout à fait fondamental pour l'analyse du débat social car ne pas parler de certains faits ou ne pas prendre en compte tel thème explicatif est tout à fait significatif d'une stratégie d'explication et je fus conduit ainsi à mettre l'accent sur le simple fait de la présence ou de l'absence. Ceci avait pour conséquence de ne pas noter l'importance

de la présence du fait qu'on ne peut pas évaluer celle de l'absence. Présence et absence furent codées pour elle- mêmes, en zéro ou un sans quantification supplémentaire.

En ce qui concerne l'identification des absences, ce problème trouva sa solution d'une manière simple : quand un thème était majoritaire, ce qui veut dire se trouvant dans plus de la moitié des journaux, ce fut son absence qui fut codée. Si sept journaux sur les onze parlent d'expédition punitive ou de ratonade à propos de ces évènements, ce sont les quatre journaux qui "oublient" d'aborder ce thème qui deviennent porteur d'information. A la limite, quand tous les journaux sans exception sont d'accord pour donner les mêmes indications de lieux et de temps dans leur description des évènements, on voit que cette question ne fait pas objet de débat et ne doit pas entrer dans l'analyse. D'une manière plus générale, on doit distinguer le "discours majoritaire" qui permet de se faire une idée sur ce qui s'est passé et le "débat d'interprétation" qui est l'objet propre de l'analyse.

### Analyse et interprétation

Le codage en présence absence conduit donc à un tableau où en ligne on trouve la centaine de thèmes issus des synopses et où en colonne se trouvent les journaux. A l'intersection des lignes et des colonnes se trouve un "1" si le thème est présent dans un journal et "0" s'il est absent, sauf dans le cas où un thème est majoritaire. Dans ce cas c'est l'absence qui est codée "1" (et la présence "0"). Sur un tel tableau l'analyse des correspondances proposa un premier facteur qui regroupait certains journaux d'une manière attendue mais d'autres d'une manière inhabituelle. Cette représentation factorielle, était doublée d'une représentation en graphe Tri-deux adaptée pour la circonstance (l'indice de Jaccard entre lignes ou entre colonnes avait été substitué à l'écart à l'indépendance).

Si certains regroupements de journaux proposés par le premier facteur étaient tout à fait compréhensibles, comme le fait que le Figaro soit proche du Parisien ou de France-Soir, ou que le Matin soit proche de Libération, on comprenait mal que le Quotidien de Paris soit proche de ce dernier groupe et non du premier. Cette situation s'expliquait si l'on regardait la représentation simultanée des thèmes : on voyait en effet que d'un côté on trouvait les thèmes qui reflétait la manière belliloise de présenter les évènements (il y avait des incidents provoqués par les jeunes et les bellilois se sont fait justice eux-mêmes) tout en en gommant l'aspect trop agressif. De l'autre, les journaux comme Libération avait pris plutôt le point de vue des jeunes et se sont rendus compte que l'affaire venait de la méprise des bellilois sur le statut des jeunes : comme ils étaient hébergés dans des locaux du Ministère de la Justice qui auparavant abritaient un ancien centre de formation de jeunes détenus (et plus anciennement encore une "colonie pénitentiaire"), les bellilois ont cru qu'ils avaient affaire à des détenus auxquels on offrait des vacances, ce qui dans un milieu où l'on est depuis plusieurs générations en rapport avec l'institution pénitentiaire, paraissait insupportable.

La découverte de cette méprise associe à cette recherche un aspect d'affaire policière élucidée qui donne de l'intérêt aux résultats de la recherche mais risque de donner de faux espoirs aux utilisateurs de la méthode. En premier lieu car les résultats peuvent être beaucoup moins spectaculaires et cependant dignes d'intérêt et en second lieu parce que l'opposition des thèmes, si elle montre bien l'articulation des thèmes autour de cette méprise, n'est certainement pas la seule méthode pour la découvrir.

Le logiciel mis au point pour cette opération fut généralisé pour le traitement des données textuelles : il permet la constitution du tableau des thèmes à partir de l'entrée des simples cooccurrences de thèmes et d'émetteurs. On inverse automatiquement les thèmes majoritaires, on fait l'analyse factorielle du tableau de présence/absence et on calcule et visualise l'indice de Jaccard entre thèmes ou entre émetteurs. Ce logiciel est utilisable

également pour des entretiens semi-directifs si une analyse préalable permet de faire émerger des thèmes par une mise en synopses : il a été diffusé à ce jour à une soixantaine de chercheurs.

Ce qui lui a été souvent reproché, c'est de laisser à la sagacité du chercheur cette étape cruciale qu'est la mise en synopses : certains, dans la lignée de l'analyse lexicologique (Lebart et Salem 1988), voudraient que l'on soit encore plus près des données en privilégiant le mot lui-même plus que le thème qui en est une interprétation par le chercheur. Dans cette logique, mais pour des raisons tout à fait contingentes, j'ai mis au point une manière de faire qui répond à ce souci : la constitution et l'analyse du "tableau lexical des questions".

## Le logiciel Tri-deux Mots

A l'origine de ce logiciel, il y eut de ma part le simple désir de rendre service aux collègues psychologues (Henri Paicheler et ses étudiants en particulier) qui appliquaient la technique de l'association de mots faite autour d'un mot stimulus. Cette technique développée assez systématiquement dans le champ des "Représentation sociales" (Cf. Doise 1992) consiste à proposer un mot stimulus et à noter les mots que l'interrogé associe spontanément à ce stimulus. En variant éventuellement les conditions expérimentales et en tenant compte des caractéristiques possédées par les émetteurs (sexe, âge, position sociale ou réponse à d'autres questions), on dispose d'une masse de données qui se révèle assez difficile à exploiter si on ne dispose que des programmes classiques d'analyse de données.

En effet, puisque l'on dispose d'une liste de mots émis en réponse et d'individus qui ont pu les émettre ensemble, le réflexe spontané est de constituer un "Tableau lexical entier" (en reprenant la terminologie de Lebart et Salem 1988) où l'on aurait la liste des mots les plus fréquents en colonne, les individus en ligne et un "1" de présence si le mot est émis par l'individu et "0" ailleurs. Ce tableau, comme Lebart et Salem l'ont également montré, se révèle peu intéressant. En effet, c'est un tableau très creux (il possède beaucoup de zéros), où les facteurs sont liés à des cooccurrences. Comme celles-ci sont nombreuses mais spécifiques, il faut de nombreux facteurs pour rendre compte des différentes associations sans qu'il y ait un facteur vraiment général. Ce peut être le salut pour un chercheur qui finira bien par y trouver l'association qui lui permettra d'aller dans le sens de ses désirs, mais on ne peut vraiment pas faire confiance aux résultats très incertains d'un tel tableau.

Par contre, si l'on croise la liste des mots associés, par importance numérique décroissante et jusqu'à une certaine limite de fréquence, avec les modalités de réponses à toutes les questions dont on dispose, on construit un tableau où l'unité statistique n'est plus l'individu, mais la cooccurrence d'un mot et d'une caractéristique sociale de celui qui l'a émis. Ce "Tableau lexical des questions", comme j'ai proposé de l'appeler (Cibois 1989b) donne immédiatement par analyse factorielle un premier facteur stable et en général tout à fait interprétable.

Quand j'ai vu se développer l'utilisation de ce logiciel dans le milieu des psychosociologues, je me suis rendu compte qu'il pouvait également servir pour les sociologues pour le traitement des questions ouvertes. En effet, Lebart et Salem proposait de croiser tout le vocabulaire associé à une question ouverte (mots outils et mots pleins) avec une question clé de l'enquête ou le pré-croisement de deux questions (sexe-âge par exemple). Or, le Tableau lexical des questions n'est qu'une généralisation de ce type de croisement et cette manière de faire est tout à fait dans la logique de la perspective de l'Analyse des données. En effet, on ne fait pas de choix préalable de la question de croisement mais on prend l'ensemble des questions disponibles, d'une manière équivalente à un tableau de Burt. De ce fait le logiciel

Tri-deux Mots peut être directement utilisé pour le traitement d'une question ouverte : il suffit pour cela de considérer le vocabulaire de la question ouverte comme ayant le même statut que le vocabulaire émis à la suite d'un stimulus. Je montrerai les résultats d'une utilisation de cette méthode dans le domaine de l'orthographe au chapitre 9.

On peut bien sûr coupler la méthode synoptique avec le tableau lexical des questions : on peut associer à des caractéristiques d'émetteurs non plus des mots mais des thèmes si ceux-ci ont été déterminés préalablement par la constitution de synopses.

### Formalisation ultérieures

Le développement de méthodes formalisées inspirées par les principes de l'Analyse des données me semble toujours prioritaire : les méthodes de traitement en analyse de contenu me semblent encore pauvres par rapport aux traitements dont on dispose sur les modalités de réponses aux questions fermées. Il ne me semble pas cependant que la solution soit dans l'empirisme qui utilise directement le mot (comme on pratique en lexicologie) car c'est vraiment appauvrir au maximum la phrase. Il ne me semble pas non plus que l'analyse puisse vraiment s'appuyer sur de pures considérations statistiques comme le fait Max Reinert (1987) qui lui part du tableau lexical entier mais qui gère l'aspect creux de ce tableau (la présence de beaucoup d'absences et de peu de présences) en faisant des classifications descendantes hiérarchiques sur les lignes du tableau.

Je penserais assez volontiers, au moins à titre d'hypothèse de travail, qu'il faille aussi formaliser les relations entre classes sémantiques dans la ligne des travaux sur la "logique naturelle" développés par Jean-Blaise Grize (Bibliographie *in* Grize 1990). Ces travaux, supposent l'utilisation des formalisations informatiques récentes comme les langages orientés objets pour pouvoir être aisément réalisables, mais ils doivent en même temps être proches des démarches empiriques des chercheurs qui savent très bien rapprocher des intitulés proches. Il faut se méfier de la perte de temps de la formalisation si elle ne doit pas conduire à une amélioration des méthodes existantes. Il y a là un chantier ouvert où je vais investir à nouveau pour une recherche demandée par la Délégation interministérielle à l'insertion des jeunes et qui a pour but de formaliser environ 300 récits d'insertion (réussie ou non) de jeunes, le but étant d'arriver à une typologie. Il y a là un matériel riche, déjà constitué, écrit en langue naturelle mais suffisamment élaboré pour permettre une formalisation adaptée.

\*\*\*

## Chapitre 6 : démographie

En mettant au point une méthode d'analyse de contenu du débat social, je me rends compte maintenant que je commençais à développer un thème de recherche qui allait prendre de plus en plus de place dans la suite. Si dans le cas de l'analyse de presse sur Belle-Ile, le contenu importait moins que la méthode, il n'en va plus de même avec les recherches ultérieures, en particulier dans le champ de la natalité dont je vais parler maintenant.

Mes contacts avec des recherches touchant le thème de la natalité en France ont des origines directement liées à la conjoncture de la fin des années 70. Pendant la campagne électorale de 1978 j'avais été surpris par le fait que tant à gauche qu'à droite tout le monde était d'accord pour dire que la France était menacée d'un péril démographique mais qu'il y

avait accord sur cet objectif minimum du renouvellement des générations. Cet accord était d'autant plus surprenant que la France se singularisait par rapport aux autres nations occidentales : la RFA ne semblait pas s'émouvoir d'avoir un nombre moyen d'enfants par femme nettement en dessous du nôtre. Aux États-Unis on se souciait même de rendre la population stationnaire en luttant contre la croissance qui continuait à faire ses effets pour des raisons d'inertie démographique. Il me semblait qu'il y avait là un débat de société qu'il fallait explorer et je résolus d'y aller voir de plus près (Cibois 1982b).

En premier lieu je m'aperçus, simplement en regardant le travail des experts qui travaillaient sur les personnes âgées, que le thème du "vieillissement de la population" était pour eux complètement idéologique : une population n'est pas un être humain. Utiliser ce thème était se servir d'une crainte devant le vieillissement individuel qui était d'ailleurs tout à fait défavorable à une politique saine des personnes âgées. Quant au problème des retraites, il était bien plus lié à la conjoncture économique qu'aux problèmes démographiques. S'il fallait s'en occuper, c'était pour répondre aux problèmes économiques immédiats, non aux prétendus problèmes de l'an 2010.

En découvrant la nature idéologique des arguments avancés, je cherchais à en découvrir l'origine et je m'aperçus qu'elle se situait dans un lieu bien précis : à l'INED.

## Généalogie de l'INED

Quand le directeur de l'INED utilise la peur du vieillissement, il ne fait que reprendre les thèmes qu'Alfred Sauvy diffusait déjà dans le premier numéro de 1946 de *Population* et lui-même ne fait que reprendre des thèmes antérieurs. Il y a là continuité entre l'INED et des institutions qui l'ont précédé : le *Fondation Alexis Carrel* pendant la guerre<sup>14</sup> mais aussi plus antérieurement le *Conseil supérieur de la natalité* et l'*Alliance nationale contre la dépopulation*. Cette continuité se retrouve également au niveau des individus qui passent des anciennes aux nouvelles institutions comme Bovérat ou Landry dont Sauvy avait été le collaborateur.

Par le biais de cette Alliance je remontais facilement jusqu'à son créateur, Jacques Bertillon, démographe et statisticien<sup>15</sup>, qui s'était aperçu du phénomène qu'on appelle aujourd'hui la transition démographique et qui voyait bien qu'elle touchait en priorité la France et non l'Allemagne. Or, nous sommes dans la dernière décennie du XIXe siècle et s'apercevoir que la France aura à sa disposition moins de soldats que l'Allemagne pour une éventuelle reconquête des territoires perdus est jugé comme dramatique. Le thème de la dénatalité s'accroche alors avec celui du nationalisme et cela d'une manière durable : il reste lié au conflit franco-allemand jusqu'à 1945, bien qu'évidemment les allemands aient eux aussi fait leur transition démographique. Il reprend de plus belle dans les années 70 avec la baisse de la natalité pour se fixer ensuite sur l'immigration, le vieillissement, les retraites.

Je montrais bien par diverses citations le lien entre l'idéologie nataliste et le thème conservateur de la lutte contre les effets pervers de la libre décision de l'individu, thème qui allait évidemment fleurir sous Vichy. Il y eu d'une certaine façon occultation du natalisme au moment de la croissance démographique d'après-guerre, mais la gauche comme la droite retrouvèrent ce thème par le biais du renouvellement des générations dans la mesure où on se servait d'un indicateur conjoncturel pour faire peur. Sur ce dernier point la querelle des indicateurs déclenchée par Hervé Le Bras dix ans plus tard allait le confirmer (Le Bras 1991, Cibois 1992d).

---

<sup>14</sup> Problème repris récemment par Alain Drouard 1992.

<sup>15</sup> Le frère du créateur de la méthode anthropométrique.

Le deuxième apport de mon texte de 1982 fut de mettre l'accent sur la nécessaire explication psycho-sociale des évolutions de la fécondité : en effet le parallélisme très fort entre les courbes de l'indicateur conjoncturel de fécondité dans divers pays ayant des cultures, des histoires et des politiques familiales différentes me fit prendre conscience que c'était la conjoncture mondiale elle-même qui avait partie liée avec l'évolution de la natalité. Pour en rendre compte je mis l'accent sur ce que j'appelais l'"idéologie du redémarrage", le fait qu'après une guerre, une crise grave, le sentiment d'échec fait que les espoirs d'améliorations se reportent sur la nouvelle génération. Qu'on retrouve là le thème théologique de l'Incarnation qui sauve du péché du Monde me semblait porteur de sens puisqu'à Noël, le cadeau fait à l'enfant est bien comme Cl. Lévi-Strauss l'a montré (1952) une offrande aux dieux dont la mystérieuse obligation est un hommage aux générations futures qui sauveront le monde de l'état abominable où nous leur laissons.

Je ne suis plus aujourd'hui persuadé de la valeur de cette explication mais par contre je vois mieux que la liaison avec la conjoncture mondiale reste forte : la crise économique révélée par les chocs pétroliers avait ses prémisses dans les années antérieures où la natalité était déjà en baisse. Comment se fait le lien, je n'en sais rien, mais il me semble qu'il y a certainement quelques rapports entre le fait des évolutions de la conjoncture économique et celui de la conjoncture démographique. Non pas en ce sens que la démographie serait expliquée par l'économie (thèses d'Easterlin ou dérivées), mais par le fait que les croissances et décroissances seraient dans les deux cas influencées par un phénomène sous-jacent qui pourrait bien avoir partie liée avec le cœur de l'éthique du capitalisme, le désir d'entreprendre. Il y a là un problème à reprendre à nouveaux frais mais ce qui reste acquis, c'est que la démographie ne saurait se suffire à elle-même, ce qui est d'ailleurs admis par beaucoup de démographes.

Mon article de 1982 quant à lui fut contesté : même à la revue *Esprit*, je dus l'adjoindre d'une postface où je disais que j'étais conscient du nécessaire droit de l'état à intervenir, ce que je niais trois lignes plus loin en disant qu'il en était intellectuellement incapable. Enfin, le rédacteur en chef, Paul Thibaud exprima son désaccord dans une préface : comme le désaccord ponctuel ne m'empêcha pas de m'exprimer j'acceptai toutes ces mises en garde. A l'INED aussi, par le biais d'une invitation de Michel-Louis Lévy, on me fit comprendre que je n'avais pas été très sérieux (et *Esprit* pas davantage) en publiant un texte sans l'avoir fait revoir par des gens compétents. Comme je restais poli et compréhensif on se montra bienveillant à mon égard. Toutes les réactions ne furent cependant pas critiques, André Burguière en particulier me proposa de collaborer à une enquête sur le même sujet qui serait basée sur un sondage fait par la *Sofres* et financé par le *Nouvel-Observateur*.

### Une enquête de la Sofres sur la natalité

L'enquête était très courte et comportait d'abord trois questions sur la natalité : si elle est jugée suffisante ; sur le rapport entre la conjoncture, la situation actuelle et le désir d'avoir des enfants ; sur les conséquences de la dénatalité. Puis était étudié l'impact de diverses mesures proposées pour lutter contre la dénatalité, depuis l'aide financière aux familles ou l'augmentation des crèches jusqu'à l'interdiction de l'avortement ou de la contraception. Enfin les modalités de l'aide de l'Etat étaient discutées en se posant la question de savoir s'il fallait tenir compte du revenu des familles.

Un article très rapide fut publié dans le *Nouvel-Observateur*<sup>16</sup> dans lequel je rappelais que la question des retraites avait plus à voir avec les problèmes du chômage ou avec celui de l'âge à la retraite qui venait d'être abaissé, qu'avec la seule dénatalité.

L'intérêt de cette collaboration avec le *Nouvel-Observateur* est qu'elle mettait à ma disposition une enquête représentative sur la question dont je fis l'analyse secondaire dans le numéro spécial *d'Informatique et Sciences Humaines* consacré à *Tri-deux*<sup>17</sup>. Je fis d'abord ressortir que le nombre d'enfants souhaités qui se situait à 2,5 était peu différents entre les diverses CSP sauf si l'on faisait intervenir la proximité politique qui entraînait un désir nataliste important essentiellement chez les cadres supérieurs.

De l'analyse globale de toutes les questions de l'enquête il ressortait que d'un côté on a le monde ouvrier où la natalité reste forte et est donc jugée suffisante, qui est centré sur le monde du travail pour les hommes (craintes pour les retraites) et où le fait d'élever des enfants veut se faire reconnaître comme un travail à domicile pour les femmes (premières mesures demandées : non pas davantage de crèches mais un salaire parental, plus d'allocations, plus de congés de maternité et moins d'impôts).

A ce monde centré sur la division sexuelle du travail, s'oppose celui des cadres supérieurs (et pour beaucoup de points des cadres moyens) où l'on refuse cette division : élever des enfants doit pour une part être pris en charge par la collectivité (crèche) de façon à ne pas empêcher l'égalisation des sexes face au travail (temps partiel). Le refus de l'ancien rôle a entraîné une abstention qui s'est traduite par une baisse notable de la natalité des cadres supérieurs. La dissonance entre idéologie et pratique se trouve réduite, pour les cadres supérieurs de droite, par une résolution symbolique à travers le souhait irréaliste d'un fort nombre d'enfants.

La question des évolutions de la natalité reste ouverte : je suis de plus en plus persuadé qu'on pourra l'éclairer en se penchant sur la manière dont les individus gèrent leur rapport au temps, que ce soit en se projetant dans le futur avec un projet d'enfant, que ce soit dans le passé en gérant leur enracinement dans une culture, thème qui va de plus en plus m'occuper et que je reprendrai au chapitre 9.

\*\*\*

## Chapitre 7 : inférence

J'avais été depuis de nombreuses années en contact avec l'équipe de H. Rouanet qui développait une méthode pour le traitement des données d'observation qui soit une extension de l'analyse de la variance. Cette technique fut couplée à l'inférence bayésienne et quand fut mis au point l'inférence sur les fréquences (Bernard 1986), je devins un propagandiste de ces méthodes dans le milieu de la sociologie (Cibois *et al.* 1988b).

En effet, si l'on est amené à faire de l'inférence c'est pour la raison fondamentale que lorsqu'on observe un résultat en sciences sociales on devrait se demander s'il est généralisable, et il est bien rare qu'on le fasse. Cependant, si cette attitude est si peu répandue dans le domaine du dépouillement d'enquête, ce ne peut être sans raison. Cela vient du fait que si on

---

<sup>16</sup> N° 949 de janvier 1983.

<sup>17</sup> N°70-71, sept.-déc. 1986.

utilise des tests statistiques classiques comme le Khi-deux, on s'aperçoit qu'il sont la plupart du temps significatifs.

Ce test du Khi-deux est en effet peu exigeant : par exemple dans le cas d'un tableau à 2 lignes et 2 colonnes le degré de liberté est de 1 et le seuil critique du Khi-deux est 3,84 (au seuil de 5%). Si l'on suppose une équi-répartition des marges sur une population totale de 100 individus on a la distribution théorique suivante sous l'hypothèse d'indépendance pour le croisement entre deux questions :

	A	non A	Total
B	25	25	50
non B	25	25	50
Total	50	50	100

Si entre la modalité A et la modalité B il y a un choix privilégié et qu'un seul individu est au-dessus de l'indépendance le Khi-deux total est alors de 0,16, s'il est de 2 de 0,64 etc..

Ecart à l'indépendance	1	2	3	4	5	6
Khi-deux du tableau	0,16	0,64	1,44	2,56	4,00	5,76

On voit que dès que l'on a un écart de 5 individus par rapport à l'effectif théorique d'indépendance, le croisement devient significatif selon le test du Khi-deux. Dès que les effectifs sont un peu nombreux, le test du Khi-deux se révèle significatif et ceci explique que dans de nombreux cas on ne se pose guère de questions sur le caractère généralisable ou non d'un résultat.

Là où le problème se complique c'est quand on se trouve face à de faibles effectifs : en effet bien souvent le test du Khi-deux devient inapplicable du fait que les effectifs théoriques sont trop faibles. Bien que la correction de Yates soit toujours possible, l'obstacle le plus important se trouve moins dans des considérations techniques que devant la crainte des chercheurs devant les petits effectifs.

### Doit-on craindre les petits effectifs ?

Il est d'usage, quand on fait de la statistique, de se méfier des petits effectifs, c'est à dire des regroupements dont l'effectif est de quelques dizaines d'individus, car on a alors le sentiment de sortir du domaine des méthodes statistiques et d'entrer dans le domaine flou et incertain des méthodes qualitatives. Cependant, on observe des pratiques sociales différentes qui consistent à tenir pour fiable un faible échantillon et nous pouvons en énumérer quelques unes :

1) le sociologue quand il pratique ce qu'il appelle l'entretien approfondi, s'affranchit de tout critère statistique. Ce qu'il demande à celui qu'il interroge c'est d'être "typique" et "normal".

"Typique" d'abord de son milieu social, de son âge, de son lieu d'habitation, de son sexe : en pratique on lui demande d'être représentatif à lui tout seul de ce que l'on attend d'une partie de population examinée par quota. Ceci veut dire que pour le sociologue, il n'y a pas de

différence de nature entre un individu caractérisé socialement, et un grand nombre d'individus ayant les mêmes caractéristiques sociales. Les seules différences sont pratiques : quand on a un individu, on peut lui poser beaucoup de questions mais on peut difficilement généraliser les résultats et inversement, quand on a beaucoup d'individus, il est beaucoup plus difficile d'avoir une information précise mais elle est plus facilement généralisable.

On recherche d'autre part un individu "normal", c'est à dire ne présentant pas de problèmes psychologiques évidents, pas d'hostilité ni de parti pris sur la question étudiée, pas de handicaps ni de particularités gênantes : on lui demande en quelque sorte ce qu'on exige d'un bon échantillon, c'est d'être facilement accessible et "sans biais".

2) l'historien a, de fait, une pratique analogue : s'il découvre les mémoires d'un humble bourgeois du 17<sup>e</sup> siècle, et si ces mémoires lui semblent dignes d'intérêt, c'est à dire "représentatifs" de la mentalité de l'époque, il se gardera bien de se poser des problèmes statistiques de représentativité et se servira de ce témoignage pour éclairer la pratique d'une classe sociale, dans une région à une époque donnée (ce qui revient à le juger représentatif d'un quota de population).

3) l'ethnologue use d'informateurs qui sont en général en petit nombre et qui lui semblent représentatifs, mais l'homme de la rue, au sens strict, c'est à dire l'homme dans la rue qui cherche son chemin, pratique de même. Pour trouver son chemin, on s'aperçoit que deux types d'"échantillonnages" sont pratiqués : si les passants sont rares, on pratique un échantillonnage au hasard (le premier venu fera l'affaire), s'ils sont nombreux, on constate que celui qui cherche son chemin recherche l'homologie sociale, c'est à dire qu'il interroge de façon privilégiée des personnes de même sexe, de même âge, de même statut social. Ceci signifie qu'il utilise pour son propre compte la méthode des quotas et faisant plus confiance à des gens proches parce qu'ils sont supposés avoir les mêmes goûts, les mêmes préoccupations, et être de ce fait plus fiables.

4) l'expérimentaliste (biologiste, psychologue) lui, a l'habitude de travailler sur quelques dizaines de cas tout au plus : la difficulté de l'expérience lui interdit d'ailleurs d'en envisager plus. Limité par force dans ses effectifs, l'expérimentaliste a vis-à-vis de ses résultats une attitude intéressante car il se pose la question du pouvoir de généralisabilité de ses résultats. C'est d'ailleurs dans ce milieu que sont nées les méthodes comme les plans d'expérience, l'analyse de la variance et plus récemment les techniques d'inférence bayésienne sur les fréquences, méthodes qui peuvent tout à fait s'appliquer sur des petits effectifs dans le cas d'enquêtes réalisées par quota.

## Les techniques d'inférence bayésiennes

"Inférer" signifie tirer des conclusions d'un fait mais ce qui spécifie ces méthodes c'est l'adjectif "bayésien" qui renvoie au R.P. Bayes qui au 18<sup>e</sup> siècle et dans la tradition de Hume<sup>18</sup>, a montré que l'on pouvait tenir compte des observations passées pour inférer la probabilité de survenue d'un évènement. Où ces techniques sont directement intéressantes aujourd'hui, c'est que les méthodes qu'elles emploient sont "exactes", c'est à dire qu'elles n'utilisent pas des lois limites, et sont donc utilisables sans difficultés sur de faibles effectifs.

Sans entrer dans le détail technique de ces méthodes pour lequel on consultera le texte de Jean-Marc Bernard (1986), le principe général en est que l'on peut, à partir d'un échantillonnage aléatoire, déterminer une distribution de probabilité sur l'ensemble des

---

<sup>18</sup> Sur ce point cf. Cléro 1988

compositions parentes possibles. On voit bien que l'on s'éloigne ainsi de la procédure des tests statistiques de signification : au vu des données on ne va pas accepter ou rejeter une hypothèse, par exemple d'indépendance dans le cas d'un tableau croisé, mais l'on va voir la confiance que l'on peut avoir à situer entre deux limites le paramètre recherché.

Comme nous utilisons des techniques bayésiennes, il faudra déterminer quelle distribution initiale on suppose : si l'on se place dans une approche décisionniste, il faudra bien sûr tenir compte des observations antérieures ou d'opinions a priori. Mais si l'on se place, comme nous le ferons, dans une perspective d'analyse inductive des données comme la définit H. Rouanet (in Lecoutre 1984), on prendra des distributions initiales qui ont le moins d'influence possible sur les résultats observés et qui traduisent notre état d'incertitude quant à la vraie valeur du paramètre. Concrètement les distributions initiales choisies seront de faible "force" (leur somme sera égale à l'unité) et elles seront réparties également entre les diverses modalités de la distribution.

La question que l'on doit se poser quand on veut appliquer ces techniques au cas d'une enquête obtenue par quota est de savoir si l'on peut considérer les individus observés comme tirés d'une manière aléatoire. Ce n'est certainement pas vrai au sens strict mais cette caractéristique n'empêche personne d'utiliser intensivement ce type de sondages. D'ailleurs, si cette méthode des quotas est valide pour des effectifs importants, on ne voit pas quel critère statistique permettrait de dire à partir de quel effectif la méthode devient invalide. C'est justement pour combler ce vide que nous proposons d'utiliser les techniques d'inférence bayésienne sur les fréquences : en faisant *comme si* le tirage était aléatoire on dispose d'un garde-fou qui va nous permettre de juger du caractère de généralisabilité d'un résultat et nous empêcher de sur-exploiter (ou de sous-exploiter) un ensemble d'observations.

J'ai expliqué aux sociologues le mode d'emploi dans un article du Bulletin de Méthodologie Sociologique (Cibois *et al.* 1988b) et dans le numéro *d'Informatique et Sciences Humaines* consacré à la méthode *Tri-deux*<sup>19</sup> en montrant comment on pouvait se poser des questions de validité de petites différences dans le cas de l'enquête sur la natalité. Enfin, un autre exemple d'utilisation a été donné à propos de problèmes liés à l'orthographe (Cf. chap. 9) où j'ai utilisé des méthodes bayésiennes pour garantir la validité d'un encadrement sur des fréquences d'omission d'accent dans des copies d'étudiants. On voyait ainsi que la pratique de l'omission reste marginale et qu'elle ne porte que sur des positions non distinctives en milieu de mot et non en fin de mot (Cibois 1992c).

\*\*\*

---

<sup>19</sup> N° 70-71, 1986

## Chapitre 8 : une science sociale synchronique

### L'article de la revue du MAUSS

"Pour une science sociale synchronique" est paru dans la *Revue du MAUSS* du deuxième trimestre 1989. Il m'a permis de faire le point d'une manière précise et de poser les jalons de l'évolution ultérieure. C'est un article "radical" en ce sens qu'il pousse des idées jusqu'à leurs conséquences pour en voir la solidité. L'articulation en est la suivante : elle repose sur l'opposition entre science et savoir.

Un savoir est vraisemblable, il est fait de connaissances empiriques, de concepts qui nous viennent de la société par le biais des idéologies, des croyances ou de savoirs antérieurs : il n'y a pas de solution de continuité entre ce "savoir" du sociologue et le travail de l'expert, ni entre celui-ci et celui du journaliste, puis avec celui de tout un chacun. C'est le fruit du travail du sociologue quand il travaille sur son terrain, pour ne pas dire sa chasse gardée, avec ses concepts qui sont des impositions de son point de vue.

Une science du social au contraire n'est pas *vraisemblable* mais *sûre* : son modèle est la phonologie, c'est à dire qu'elle cherche à trouver la règle du jeu des oppositions significatives, des traits pertinents à un moment donné. Elle ne peut se constituer que par observation de la manière dont des traits sociaux de situation, de comportement, d'opinion, classent les individus dans des groupes.

Mon projet durkheimien arrive ici à son acmé : il se radicalise. La scientificité du structuralisme linguistique en est le ressort : la compréhension wébérienne est rejetée dans le vraisemblable du savoir, utile certes pour la vie courante, pour la politique, pour l'utilité sociale, mais la barre de la scientificité est placée au-dessus. Projet durkheimien mais très critique de Durkheim dont plusieurs ambiguïtés sont relevées : en particulier est bien montrée la liaison entre son souci de lutter contre les prénotions et son idéologie sociale très datée qui voyait dans la corporation le creuset social qui allait remplacer une religion respectée mais éliminée comme source du lien social.

Un point qui se révélera essentiel est ici bien mis en lumière : il faut soigneusement distinguer la scientificité qui ne peut être que synchronique, être une règle de grammaire, une règle du jeu, une prise en compte des lois d'une harmonie à un instant donné, en un mot un jeu d'opposition ; il faut distinguer cette scientificité du *savoir social* qui porte sur les évolutions de la société, qui s'enracine dans l'histoire, la tradition, les idéologies, les croyances. Il n'y pas de science de l'évolution, mais il n'y a pas non plus de "compréhension" des systèmes d'oppositions. On ne peut pas mettre d'idéologie sur une grammaire ni de rationalité sur l'histoire. Il y a un croisement de deux directions qui sont irréductibles l'une à l'autre.

On retrouve là l'acquis de la philosophie : P. Ricoeur (1983) nous montre que l'histoire est "mise en intrigue" rappelant lui-même Saint-Augustin qui, dans les *Confessions*, tente une première conceptualisation du temps. Savoir ce qu'est le temps, c'est rationaliser la succession et cela se révèle impossible : on ne peut que constater la présence du passé et celle du futur.

Inversement on retrouve l'acquis de l'épistémologie des sciences : une science "normale" au sens de Kuhn, c'est à un moment donné, un ensemble de concepts qui forment système. S'il y a évolution, c'est par changement de paradigme, de système, par rupture brusque sans progressivité parce que les jeux d'oppositions forment un tout cohérent. On peut briser le système, non le modifier.

Comme cet article ne posait que des principes, il était indispensable de tenter de les mettre en œuvre à travers une réalisation concrète dont je pourrais tirer des leçons. L'appel d'offre pour une réexploitation de l'enquête sur les pratiques culturelles des français de 1989 m'en fournit l'occasion (Ministère de la culture 1990).

## Les pratiques culturelles des français

Il s'agissait de mettre en œuvre un principe théorique sur des données concrètes qui n'avaient pas été conçues à cette fin mais pour tester l'avancée ou le recul de pratiques culturelles. Cependant, comme on avait voulu éviter les effets d'une bonne volonté culturelle, le questionnaire était présenté aux 5000 interviewés comme une enquête sur les loisirs, ce qu'elle était en partie d'ailleurs.

Mon premier travail consista à critiquer l'analyse factorielle publiée dans le rapport du ministère (Donnat et Cogneau 1990) et qui prétendait montrer que la principale opposition était faite par des activités "branchées" très minoritaires. Je découvris à cette occasion l'existence d'un "effet de distinction" en analyse des correspondances qui vient du fait que s'il existe dans une population étudiée un sous- groupe qui recherche des pratiques distinctives, ce sont ces pratiques qui créeront obligatoirement l'opposition du premier facteur du fait de l'utilisation de la distance du Khi-deux qui met en relief les attractions liées à de faibles effectifs (Cibois 1992a).

Je mis donc au point une analyse non pondérée qui, sur le même objet pertinent que l'analyse des correspondances, les écarts à l'indépendance d'un tableau de Burt, utilise une distance euclidienne ordinaire. Cela revient à faire de l'analyse en composantes principales sur un tableau centré mais non réduit. Je pus ainsi faire une analyse d'ensemble qui montra que les oppositions majeures étaient le fait du niveau socio-culturel (repéré par le niveau d'étude BAC ou plus, le fait d'être ouvrier ou le fait de ne pas avoir de diplôme ou le CEP certificat d'études primaires), la situation dans le cycle de vie (célibataire, marié, retraité) et le sexe. Je croisai l'ensemble des pratiques de l'enquête avec ces huit caractéristiques fondamentales : je regardai pour chacune l'intensité de la liaison entre modalités de pratique et modalités de statut.

## Styles de vie

En croisant les huit modalités pertinentes de statut avec les pratiques culturelles ou de loisir, on obtient pour chaque cas une "signature", c'est à dire un ensemble de huit valeurs du PEM. On peut de ce fait repérer des similitudes entre signatures, par exemple entre deux émissions de télévisions que sont *Des chiffres et des lettres* et *Les enquêtes du commissaire Maigret* qui ne sont pas du tout de même genre mais qui, on le voit sur cet exemple, sont suivies par des gens aux caractéristiques proches :

	Niveau social			Position			Sexe	
	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
Maigret	74	91	118	47	95	117	98	102
Chiffres et lettres	54	64	134	57	70	133	98	102

Les PEM sont ici centrés sur 100 : un PEM de +17 est égal à  $100 + 17 = 117$ , un PEM de -26 est égal à  $100 - 26 = 74$ , 100 correspond donc à l'indépendance.

BAC= Bac ou plus, OUV= ouvrier, CEP= pas de diplôme ou Certificat d'études

CEL= célibataire, MAR= marié, RTR= retraité

MAS= Masculin, FEM= Féminin.

On constate que les deux signatures sont très proches : même rejet des diplômés et des plus jeunes, même choix des retraités peu diplômés. Quant au sexe, même petite différence pour les femmes.

Une fois que nous avons traité l'ensemble et malgré tous nos choix préalables, nous nous retrouvons dans une perspective (phon)étique : nous sommes submergés par une multitude de chiffres si nous envisageons de comparer les environ 200 modalités de comportement que nous avons retenues. Plusieurs solutions peuvent être envisagées qui sont la classification automatique et la simplification des données pour trouver un jeu d'oppositions.

#### 1) la classification automatique

Nous restons dans une perspective "étique" puisque nous prétendons travailler sur les données dans toute la richesse de la signature. L'expérience faite avec deux types de distances (euclidienne, khi-deux) et quatre types d'agrégation (saut minimal, distance moyenne, variance, barycentre) conduit à une grande diversité de résultats. Les seules constantes sont celles qui rapprochent des modalités très semblables comme les deux que nous avons présentées plus haut.

Les classifications, si on les rapproche, donnent des résultats non comparables : comme on n'a aucune raison de prendre une option plutôt qu'une autre, on doit abandonner cette méthode.

#### 2) la simplification des données

"Qui perd gagne" dit la sagesse populaire (peut-être pour se consoler de ses échecs), et nous sommes amenés à simplifier les signatures en les regardant d'un point de vue sociologique. Si on regarde par exemple les deux émissions citées plus haut, on s'aperçoit qu'elles sont socialement "marquées" par l'âge et le faible niveau de diplôme. Par contre l'émission *Apostrophes* est surtout marquée par les "Bac et plus" (comme *Océaniques*), tandis que *Ushuaia* et *Musicales* le sont par les jeunes. De même on pourra mettre ensemble *Maguy*, *Avis de recherche*, *La roue de la fortune* et *Barbara* qui sont spécifiques d'un faible niveau de diplôme et du sexe féminin.

Pour simplifier l'information, on n'en retiendra que la présence ou l'absence de ces "marques" sociales, le fait qu'il y a un niveau d'attraction suffisant (en général un PEM de +10%) entre une émission et une caractéristique de statut. Avec cette méthode on peut ainsi classer d'une manière lexicographique les 200 profils dans la mesure où seront semblables les comportements qui sont en attraction positive suffisante avec des modalités de statut. Soit par exemple les deux premiers profils qui correspondent aux jeunes (CEL) diplômés (BAC) soit de sexe masculin soit de sexe féminin. On s'aperçoit que les garçons jouent au tennis tandis que les filles ont des activités artistiques, que les garçons lisent des quotidiens ou des livres de sciences et techniques, des magazines, des BD ou de la science-fiction tandis que les filles lisent de la littérature classique, que les garçons vont au concert de jazz ou de rock tandis que les filles vont au spectacle de danse ou à des expositions de peinture, que les garçons vont au cinéma (sans autre indication), tandis que les filles y vont voir des films d'auteurs.

Séréotypes que toutes ces oppositions ? Il se peut en effet que par le biais d'une enquête on ne fasse que retrouver des rôles intériorisés, mais ce que l'on oublie c'est qu'il s'agit des

stéréotypes *en milieu jeune et diplômé*. Si l'on prend la même opposition dans d'autres contextes de statut, on rencontre d'autres réalités. Les exemples sont innombrables :

- toujours diplômés mais non plus jeunes mais mariés : la responsabilité d'organisation est masculine, le fait de lire des romans contemporains est féminin.

- toujours diplômés mais retraités et donc plus âgés : les hommes lisent des livres sur l'histoire ou regardent les émissions d'A. Decaux tandis que les femmes disent préférer les romans historiques. En matière de télévision les hommes choisissent *L'heure de vérité* plus politique tandis que le *Grand échiquier* plus culturel est préféré par les femmes.

Evident, toujours évident peut-on dire, mais qui l'aurait dit avant qu'il y avait là opposition de sexe dans un contexte bien particulier d'âge et de niveau social ?

- en milieu ouvrier, célibataire ou marié, l'homme dit regarder *Ushuaia*, la femme dit suivre des cours pour améliorer sa situation professionnelle. Cela veut-il dire que la condition ouvrière est vécue par l'homme sous le mode de la fuite par l'évasion et par la femme sous le mode de la sortie par la promotion ? Comme ce n'est plus évident, cela va être contesté. Prouvez-le dira-t-on. C'est une hypothèse d'interprétation, ce qui est sûr, c'est l'opposition elle-même car elle n'était pas recherchée, c'est le sous-produit non voulu d'une démarche formelle.

- chez des gens non diplômés, mariés ou retraités, les hommes font un jardin potager, les femmes un jardin d'agrément. Ce qui est intéressant ici, ce n'est pas l'opposition de sexe, mais la caractéristique générale du milieu qui fait du jardin : populaire et déjà d'un certain âge. D'autre part les hommes regardent des westerns et les femmes *Au théâtre ce soir*.

- question au lecteur : aimer les films de guerre ou historiques est le fait de qui ?

Réponse : d'hommes retraités sans diplôme qui disent aussi aimer se reposer sans rien faire de précis tandis que les femmes de même situations disent faire de la broderie, du crochet ou de la tapisserie. Le retraité sans diplôme peut avouer ne rien faire, sa femme non. Par contre le retraité masculin diplômé partage avec le non-diplômé son goût pour l'histoire mais il passe par l'intermédiaire de livre ou de télévision.

Étudions maintenant des comparaisons de niveau social par exemple en prenant des jeunes de sexe masculin et en faisant porter l'opposition sur le fait d'être "bac ou plus" contre le fait d'être ouvrier

- tous aiment le rock, mais les diplômés vont au concert tandis que les ouvriers achètent des disques. Opposition financière ? Mais faire du jogging peut ne pas entraîner de dépenses aux diplômés tandis qu'aimer les jeux électroniques peut coûter une fortune aux jeunes ouvriers. Quant à la science-fiction aimée des deux groupes, elle passe par le roman pour les diplômés et par le film pour les ouvriers. Différence naturelle de niveau d'étude, certes, mais le bridge est bien côté culturellement et le tarot également chez les jeunes. Pourquoi seuls les ouvriers déclarent aimer jouer aux cartes ?

Arrêtons là cette énumération puisque je n'ai pas (encore) trouvé un système convaincant d'oppositions qui la résume et en montre la logique. Si on en cherche une vision simplifiée, il suffit de reprendre l'analyse générale qui en donne un bon plan factoriel. La jeunesse s'allie au niveau de diplôme et la vieillesse à l'absence de diplôme (évidemment pour des raisons historiques). On peut passer d'un pôle à l'autre soit par la culture ouvrière, soit par la culture diplômée. Quant à l'opposition des sexes elle ne peut apparaître qu'en troisième facteur car elle est présente partout. De même le fait d'être marié introduit dans tous les contextes une transition entre les jeunes et les âgés.

Cette vision simplifiée est simpliste et peu intéressante. Ce qui fait l'intérêt de la formalisation effectuée, ce passage du (phon)étique au (phon)émique (ou phonologique), c'est :

1) la possibilité offerte de caractériser un ensemble de modalités de statut par un ensemble de comportements, de donner ainsi un statut formel au concept si galvaudé de "style de vie" (et critiqué justement, cf. Herpin 1986) : dire que le style de vie de la femme d'un certain âge (mariée ou retraitée) sans diplôme c'est de recevoir à diner ses parents, de faire du jardinage, de tricoter et de regarder Au théâtre ce soir, tandis que celui de l'ouvrière mariée est de rechercher de nouvelles recettes de cuisine dans des livres et de lire des romans Harlequin ; c'est donner ainsi une précision aux styles de vie qui me semble évidemment plus intéressante que de parler de recentrés, de décalés ou de vigiles.

2) la possibilité de retrouver des oppositions stéréotypés (de sexe, d'âge ou de niveau d'études), mais en voyant le contexte social, parfaitement délimité, où elles se situent.

A cela on peut répondre qu'il existe des marquages qui dépassent les contextes : que le livre, la cuisine, la couture sont marqués au féminin, que la voiture, la technique, la force physique sont marquées au masculin. C'est quelquefois faux donc sans aucune utilité : ce sont les garçons diplômés qui évoquent les surgelés : ce sont les hommes diplômés qui parlent d'autres lectures que de romans ; ce sont les femmes diplômées qui parlent de gymnastique. On pourra trouver des explications pour chacun des cas cités : on peut toujours rationaliser, pêcher à la ligne des bonnes raisons : ce dont on a besoin c'est d'un cadre général qui permette de situer avec précision chaque observation.

### Aller plus loin ?

On peut souhaiter aller plus loin en approfondissant la notion de système général d'opposition dont on a présenté ici une esquisse. Si nous ne sommes pas arrivé à ce système général, c'est certainement par manque d'imagination mais aussi peut-être parce que les données n'ont pas été constituées dans ce but. Il faudrait refaire une vaste enquête où les différentes pratiques soient bien différenciées : où par exemple on ne mélange pas le bridge, le tarot et la belote sous la seule rubrique "jeu de cartes", où de même on distingue entre les divers "sports collectifs" car foot, hand, rugby ou basket n'ont pas la même connotation. Il faudrait de plus trouver un répertoire d'activités et de comportements qui soit aussi détaillé selon les milieux sociaux, les âges ou les sexes alors que l'enquête sur les Pratiques culturelles des français était implicitement centrée sur la "vraie" culture, celle qui est plutôt pratiquée par les milieux situés en haut de l'échelle sociale.

\*\*\*

## Chapitre 9 : sociologie de la culture

Je voudrais dans cette dernière partie rendre compte de certaines de mes recherches actuelles que je mets sous la rubrique de la sociologie de la culture faute d'un terme plus adéquat. Il s'agit d'une part des recherches sur l'exploration du débat social et la natalité déjà exposées et d'autre part d'investigations sur la réforme de l'orthographe, sur l'enseignement du latin et sur un courant pédagogique réformateur appelé *l'Education nouvelle*. Je dois dire que l'unité de ces différentes recherches ne m'est pas apparue immédiatement mais que j'en vois mieux aujourd'hui la logique interne. Elle vient de ce que dans ces différents domaines, j'étudie sociologiquement la manière dont est géré le rapport au passé, en particulier dans ses composantes culturelles. La question que je me pose est de savoir comment les individus acceptent ou refusent des modifications de certains éléments culturels qui leur viennent de leur histoire personnelle ou collective et comment ils recomposent ces éléments quand ils les modifient.

### I La réforme de l'orthographe

Que les lettres de l'alphabet latin soient mal adaptées au système phonologique du français se voit immédiatement : on n'y dispose pas de la distinction entre le "i" et le "j", ni de celle entre le "u" et le "v", lettres différentes que Corneille a imposé le premier dans les éditions de son théâtre alors qu'elles n'existaient pas en français et qu'on a dû utiliser pour les différencier les oppositions entre majuscule et minuscule du latin (Cibois 1989d). On n'y dispose pas non plus de la distinction entre le "e muet", le "e ouvert" et le "e fermé" : on a d'abord utilisé des lettres muettes "z" ou "s" (qui subsistent dans est, les, chez, nez) avant de recourir au 18<sup>e</sup> siècle au système des accents. Quant aux nasales, il fallut utiliser des lettres muettes ("dans", "en", "emporter") ou des redoublements comme dans "ennui" (ou même dans "année" dont la première syllabe s'entend encore nasalisée dans l'est et le sud de la France).

Le système orthographique français est rendu très difficile du fait :

- de la subsistance de lettres mortes dont on ne sait plus s'il faut les prononcer ou non (dans "compteur" le "p" est bien senti comme muet alors que dans "dompteur" il ne l'est plus pour certains : la distinction devient sociale),

- de la présence simultanée de plusieurs manières graphiques d'arriver à la même prononciation : pour exprimer le son "e ouvert", on peut soit mettre un accent grave comme dans "je cède", soit redoubler la consonne comme dans "je jette", soit utiliser un accent circonflexe comme dans "extrême",

- de la présence simultanée d'évolutions graphiques historiques différentes. Le "s muet" de "boscage" a disparu dans "bocage" (cf. "bosquet") mais pas dans "hôte" (alors qu'il a disparu de "otage" de même racine). L'accent circonflexe n'a pas disparu de "coût" (Cf. l'anglais "cost" de même racine) alors qu'il a disparu de "coutume" (Cf. "costume" ou l'anglais "custom" de même racine). Pourquoi cet accent circonflexe se maintient quelquefois comme trace d'une lettre muette alors qu'il a disparu de "soutenir, plupart, coteau, pacage, levure, masse, goitre, gaine" ?

Dès sa création, l'Académie française a cherché à rationaliser le système : avec cinq réformes au 18<sup>e</sup> siècle, elle a pu passer du français de Corneille :

"Rodrigue qui l'eu dist, Chimène qui l'eu crust"

à celui de la *Déclaration des droits de l'homme et du citoyen* où sur un texte assez long les deux seuls mots qui ne s'écrivent pas comme aujourd'hui sont les "représentans du peuple françois".

La Révolution n'amplifia pas les rationalisations que le siècle des Lumières dont l'Académie (avec Voltaire) était un bon reflet, mais à partir du moment où la République s'installa, l'Académie française devint un des pôles de résistance du légitimisme et logiquement refusa de continuer le travail de l'Académie des Lumières. Par contre l'Académie depuis la III<sup>e</sup> République se refuse à des modifications jusqu'à une date récente sous la pression des réformateurs.

En effet, les hésitations du système entraînent celles de celui qui écrit et donc ses fautes d'orthographe. Elles ne sont pas faites au hasard mais dans les secteurs les plus "dangereux" que sont les accents et les redoublements. A partir du moment où il fallut faire un enseignement de masse de l'orthographe française, des enseignants protestent et demandent des simplifications du système. Au début du siècle, avec l'appui d'Anatole France, des tentatives sont proposées dans l'enseignement par le biais d'un arrêté de 1901 vite caduc. De nombreuses tentatives ont eu lieu depuis lors jusqu'à ces années dernières où j'ai personnellement pris une part active aux tentatives et où j'ai beaucoup appris.

Le fait de tenter une action sociale fait apparaître les appuis, les oppositions, met à jour le fonctionnement de la société sur ce point précis. Cet aspect révélateur de l'action m'a permis par l'utilisation des enquêtes disponibles de me faire une idée assez précise du fonctionnement social en matière d'orthographe (Cibois 1989b, 1989d, 1991).

### Le paradoxe enracinement/fonctionnalité

Une seule enquête rendue publique a été faite sur un échantillon national : il s'agit de celle de la revue *Lire* (*Lire* 1989) qui date de mars 1989, c'est à dire à un moment où le problème de la réforme est déjà présent, mais où le gouvernement n'a pas encore décidé de mettre en place une commission ad hoc.

A la question de savoir si l'on est favorable ou non à une réforme de l'orthographe, 44% répondent qu'ils sont favorables, 50% qu'ils sont hostiles et 6% ne se prononcent pas. Il y a donc bien une division en deux blocs de la société sur cette question, même si les avis les plus extrêmes (très favorables 12% et très hostiles 25%) ne regroupent qu'un tiers de la population. Si l'on veut en savoir plus sur cette coupure, c'est à dire mettre en rapport les réponses aux autres questions des gens favorables à la réforme et celles des hostiles, on est obligé de recourir à d'autres enquêtes, car si les résultats globaux de l'enquête de *Lire* sont publics, les données n'ont pas été communiquées pour une analyse secondaire. Il faut donc s'en tenir à des échantillons partiels qui nous éclaireront cependant car on y retrouve toujours la même structure, ce qui plaide en faveur de la typicalité de ces échantillons même sans garantie de représentativité statistique.

Dix-sept personnalités ont été interrogées par la revue *Lire* selon le même questionnaire que l'échantillon national, mais les réponses individuelles ont été publiées ce qui permet de faire une analyse globale des réponses (Cibois 1989d : 116-120) et de constater que du côté des "hostiles" on trouve l'orthographe facile, qu'il serait impossible de réformer l'orthographe sans dénaturer la langue, que la grammaire est facile, que c'est un charme de la langue, qu'il n'est pas urgent de simplifier l'orthographe pour faciliter l'apprentissage de la langue française. On est d'autre part opposé à toutes les réformes envisagées dans le questionnaire : suppressions des accents circonflexes, des traits d'union, des consonnes doubles. On est opposé au remplacement de ph par f et à celui de x par s dans les pluriels en -oux. En un mot

ceux qui sont hostiles à la réforme, le sont globalement et n'en voient pas la nécessité car pour eux il n'y a pas de problème de l'orthographe.

De l'autre côté, se trouvent de manière symétrique ceux qui trouvent que l'orthographe est difficile et qui sont favorables à une réforme. Cependant ce groupe se scinde en deux sous-groupes : celui des réformateurs engagés, favorables aux diverses modifications proposées, qui pensent que la réforme est urgente et qu'elle doit être initiée par une instance collective : dans ce groupe se trouvent Michel Rocard, Serge July et André Fontaine, ce qui est tout à fait intéressant pour la suite des événements. L'autre sous-groupe est beaucoup plus modéré et ne veut qu'une réforme minimale.

Au centre de tout, proche de la moyenne nationale, on trouve les opinions suivantes qui sont donc partagées par les favorables et les hostiles à l'orthographe : l'orthographe fait partie de notre patrimoine (86% le pensent dans l'échantillon national), elle est difficile (73%) mais pas un casse-tête inutile (83%), en matière de réforme on pourrait envisager d'en retoucher quelques bizarreries (76%). La présence de ce "centre" où se trouvent des réponses par ailleurs massives au plan national manifeste la présence d'un socle commun à tous, à base de respect vis à vis de l'orthographe mais en même temps de prise de conscience de ses difficultés et des améliorations marginales qui peuvent être faites.

Une deuxième enquête, réalisée en novembre 1989 (donc dans le même contexte que l'enquête précédente, Cibois 1989b) auprès d'une centaine d'étudiants, va nous permettre de mieux comprendre certains aspects de la perception de la réforme. En effet on a fait passer le même questionnaire que celui de l'enquête précédente de *Lire* mais en rajoutant une question ouverte : "indiquez en quelques phrases ce que vous pensez de la réforme de l'orthographe". On a ensuite constitué un tableau où à chaque ligne correspond un mot du vocabulaire de fréquence 4 et supérieure et où à chaque colonne correspond une modalité de réponse aux questions fermées. A l'intersection d'une ligne et d'une colonne se trouve le nombre de fois où un mot est émis par un individu ayant la caractéristique de la colonne. On fait ensuite l'analyse des correspondances d'un tel tableau ce qui permet d'éclairer les mots par les caractéristiques de ceux qui les ont émis (Technique évoquée au chapitre 5).

Si l'opposition la plus massive se fait toujours entre ceux qui trouvent l'orthographe facile et qui sont hostiles à toute réforme, et ceux qui la trouvent difficile et qui sont favorables à des réformes, on constate qu'il existe un deuxième principe d'opposition qui vient diviser ces deux camps. Par exemple dans le groupe des hostiles on voit apparaître une distinction entre les "très hostiles" qui sont mêmes opposés à la correction des bizarreries de l'orthographe, et les "assez hostiles" qui paradoxalement acceptent de confier la réforme de l'orthographe à diverses instances comme l'Académie française ou les ministères de la Culture ou de l'Education nationale.

Cette contradiction entre une hostilité à la réforme et le fait de la confier à une instance n'est qu'apparente et est expliquée si l'on tient compte du vocabulaire associé plus souvent que la moyenne à ces réponses : on y fait mention en effet de "certaines", "modifications", "simplifier", "changement", "difficulté", "apprentissage", "enseignement". On est donc hostile à la "réforme", mais pas à des "modifications" et le reste de l'enquête manifeste bien qu'à l'idée de "réforme" est associées l'idée de réforme *radicale*, phonétique, utopique, impossible, même quand on la souhaite ; et qu'il faut lui opposer l'idée de modification ponctuelle, limitée, raisonnable à laquelle on peut être favorable, même si on est contre la "réforme".

De ce fait on comprend mieux pourquoi le Premier ministre dans son discours du 24 octobre 1989 installant le Conseil supérieur de la langue française éprouve le besoin de se démarquer de l'idée de réforme étant donnée la connotation radicale qui lui est associées :

"Une véritable réforme, qui modifierait les principes mêmes de la graphie de notre langue, et altérerait donc son visage familier, me paraît absolument exclue. A vrai dire, personne ou presque ne la propose. (...) En revanche, en dépit des modifications effectuées depuis deux siècles, il reste encore à opérer des "rectifications utiles"". Ceux qui sont favorables, comme le Premier ministre, à un réajustement limité ont intérêt à abandonner le terme de "réforme". C'est donc ce qui fut fait et le Conseil supérieur de la langue française publié au JO en décembre 1990 porte dans son titre le terme de "rectifications".

Ce rapport eut un important impact dans l'opinion et modifia la problématique antérieure dans la mesure où, reprenant à quelques nuances près le programme de l'association AIROE<sup>20</sup>, il abandonne les projets de réforme de la commission Beslais<sup>21</sup> et se donne des objectifs bien plus limités : régularisation des traits d'union, du pluriel des mots composés, de diverses anomalies ; suppression de l'accent circonflexe sur le i et sur le u ; modification du participe passé des verbes pronominaux.

Afin de juger ces "modifications" proposées qui ont changé la nature du débat, une enquête a été faite en octobre 1990 auprès d'une centaine d'étudiants (de premier cycle et deuxième cycle universitaire parisien de sciences humaines) avec un nouveau questionnaire en 5 parties :

1) reprise de quelques questions de l'enquête de *Lire* à des fins comparatives : trouvez-vous l'orthographe facile ou difficile ; seriez-vous favorable ou hostile à des modifications de l'orthographe ; pour vous l'orthographe est-elle une difficulté inutile ; quelque chose qui fait partie de notre culture, de notre patrimoine.

2) dans une deuxième partie, on reprend les modifications proposées par le *Conseil supérieur de la langue française* : suppression de bizarreries, soudure de mots composés, simplification de pluriels de mots composés, suppression de l'accent circonflexe dur des i ou des u. On demande si pour chacune de ces modifications on est favorable ou non.

3) on reprend ensuite, en signalant qu'elles n'ont pas été retenues, les propositions anciennes de réforme : passage du x au s dans les pluriels en -oux ; suppression des lettres grecques ; des consonnes doubles ; invariabilité du participe passé conjugué avec "avoir". On teste l'accord éventuel avec ces modifications.

4) on propose ensuite un certain nombre d'opinions relatives à l'orthographe que l'on demande d'approuver ou non : opinions relatives à l'urgence des simplifications pour faciliter l'apprentissage, à la possibilité de supprimer les bizarreries, au problème du bon ou du mauvais enseignement ; à l'effet supposé de la méthode globale ; de la télévision ; du goût à la lecture ; de l'importance donnée à la Culture ; à la politesse ; à la politique.

5) enfin on propose une question ouverte en demandant d'indiquer en quelques phrases ce que l'on pense de l'orthographe et de ses modifications éventuelles.

En essayant de croiser le vocabulaire de la question ouverte avec les réponses aux questions fermées, on constata un éparpillement du vocabulaire qui interdisait une interprétation cohérente et l'on fut donc amené à employer une méthode d'analyse de contenu pour condenser l'information. J'ai repris à cette fin la méthode "synoptique" mise au point dans l'examen d'un corpus de presse (Cf. chap.5, Cibois 1985) et dont le point de départ théorique est de ne pas rechercher des thèmes a priori, mais de les laisser émaner du corpus

---

<sup>20</sup> Association pour l'Information et la Recherche sur les Orthographe et les systèmes d'écriture, groupe de pression formé d'enseignants et de linguistes, très actif sur le problème de la réforme.

<sup>21</sup> Commission mise en place par le Général de Gaulle à la suite d'un vœu de l'Académie des sciences : les réformes proposées qui étaient assez ambitieuses furent bloquées par le Président G. Pompidou

lui-même. Pour ce faire on met en synopse des éléments de contexte proches que l'on peut rassembler sous un même thème simplement en opérant les opérations de standardisation de vocabulaire ou de résumé (Cibois 1991).

En appliquant cette méthode à l'ensemble des textes des 110 répondants qui ont émis un texte, on arrive à découper l'ensemble en 236 éléments de contexte correspondants chacun à un thème. Quant aux éléments de contexte, on peut les regrouper en 17 thèmes qui permettent d'en classer 91%, 17 thèmes que nous allons répartir en trois grandes classes.

#### 1) Les thèmes d'enracinement

Sous cette rubrique on regroupe les thèmes qui mettent l'accent sur le fait que l'orthographe est enracinée dans le passé, que ces racines soient étymologiques ou historiques : à cet enracinement est associé une culture à laquelle beaucoup vouent un attachement très fort, perçu d'un point de vue esthétique. L'orthographe est sentie comme ayant sa beauté, son charme propre dans sa complexité même. Voici quelques exemples typiques de ces thèmes :

##### Etymologie :

- l'orthographe actuelle trouvant ses racines dans le latin et le grec il ne faut pas la modifier sous peine de perdre de nombreux points de repères permettant parfois de trouver le sens de mots mal définis

##### Histoire :

- il est un héritage de très longue date et renferme des siècles d'histoire : l'orthographe de certains mots ou expressions renvoie à des phénomènes sociaux du passé d'où l'utilité de ne pas tout bouleverser

##### Culture :

- l'orthographe et la culture des mots (poésie, littérature) tombent en désuétude et c'est bien dommage car cela fait partie intégrante de notre patrimoine culturel

##### Beauté :

- je suis pour qu'on conserve les accents circonflexes je trouve cela tellement esthétique

##### Charme :

- cependant c'est bien dans cette difficulté que réside son charme

#### 2) Les thèmes fonctionnels

On place dans cette catégorie les thèmes qui insistent sur le fait que l'orthographe est un outil qui doit donc être adapté à sa tâche : on y trouve les thèmes qui soulignent les bizarreries, les difficultés qui gênent l'utilisateur ; les simplifications indispensables pour que le plus grand nombre puisse avoir accès à l'écrit.

##### Bizarreries :

- cependant je pense que ces modifications ne doivent concerner que les absurdités et les bizarreries de la langue française

##### Difficultés :

- la langue française est indubitablement une des plus difficiles au niveau de l'orthographe

##### Simplifications :

- il est vrai qu'on peut peut-être simplifier quelques manières d'écrire les mots

Même la grammaire pourrait être modifiée :

- il y a effectivement certaines règles grammaticales qui peuvent être modifiées mais dans la mesure où elles se révèlent être complètement inutiles

Accès du plus grand nombre :

- rendre l'orthographe plus accessible c'est faciliter l'accès de plus de personnes à la langue à la lecture

3) Les thèmes de controverse

Comme nous l'avons vu dans l'échantillon national, l'enracinement de l'orthographe (c'est un patrimoine) est perçu comme quelque chose d'essentiel par la grande majorité, comme de même le thème fonctionnel lié à la suppression des bizarreries. Ceci rend le débat très controversé dans la mesure où si l'on insiste sur un thème, on en vient à négliger l'autre : si l'on met l'accent sur l'enracinement historique de l'orthographe qu'il faut préserver, on rend plus difficile son adaptation comme outil fonctionnel. Inversement, si l'on veut qu'il soit plus simple, adapté au besoin du plus grand nombre, on en vient à opérer des simplifications qui touchent à l'acquis historique. De ce fait, à la croisée de la dimension historique et de la dimension synchronique et utilitaire, on trouve ce que j'appelle les thèmes de controverse, c'est à dire les thèmes qui soit, expliquent pourquoi l'on doit réformer malgré telle difficulté ou au contraire, pourquoi il faut s'abstenir de réformer.

Thèmes en faveur de la réforme

Modifications :

- dans une perspective d'égalisation des chances devant l'institution scolaire certaines réformes concernant notamment les exceptions peuvent être entamées

Evolution :

- à bas l'illogisme : l'orthographe doit évoluer. Heureusement réforme ou pas l'orthographe à toujours évolué d'une façon officieuse

Thèmes contre la réforme

On doit conserver l'orthographe comme aide à l'apprentissage de la logique :

- l'orthographe est aussi un moyen d'analyse logique et d'apprentissage de cette logique pour les enfants

L'orthographe est mal enseignée :

- le problème posé ici n'est pas celui de l'orthographe par elle-même mais de l'éducation scolaire et du niveau de recrutement des élèves dans les classes : doit-on plutôt simplifier celle-ci ou améliorer aider les connaissances de chacun ?

On ne réforme pas l'orthographe par décret :

- je pense que personne n'a le droit ni ne doit avoir le pouvoir d'instituer une orthographe modèle

C'est un problème de laxisme :

- à ce rythme on pourrait aussi proposer des réformes en mathématiques et dans d'autres matières

Difficulté de la cohabitation entre nouvelle et ancienne orthographe :

- cependant problème pour les générations qui ont appris une orthographe classique face aux changements

## Situation du débat

Si maintenant on croise les modalités de réponse aux questions fermées avec les thèmes mis au point, l'analyse permet de retrouver l'opposition fondamentale entre partisans et adversaires de la réforme. Du côté des adversaires on pense qu'il ne faut pas supprimer les bizarreries, on est hostile à des modifications, on pense que l'orthographe est mal enseignée, que la télévision est fautive, les parents également, qu'il n'y a pas d'urgence à modifier, que l'orthographe n'est pas une difficulté inutile. Du côté des partisans, on est favorable aux modifications proposées, qui sont urgentes car l'orthographe est une difficulté inutile qui ne fait pas partie de notre patrimoine.

L'intérêt de l'analyse ne réside pas dans cette opposition, qui est permanente, mais dans la manière dont les thèmes viennent s'y associer. En effet on constate que sur les 17 thèmes, 6 à eux seuls apportent 91% de la contribution au premier facteur de l'analyse des correspondances du tableau<sup>22</sup>. Du côté hostile aux modifications on trouve les trois thèmes "laxisme", "mauvais enseignement", "orthographe élément de notre culture" ; et du côté favorable aux modifications, les trois thèmes "faciliter l'accès à l'écriture pour le plus grand nombre", "évolution possible", "simplification nécessaire". Les autres thèmes sont au centre, sans opposition sur le deuxième facteur qui semble être lié à des effectifs extrêmement faibles.

Ceci peut être interprété comme manifestant que l'opposition fondamentale n'est pas entre l'enracinement et la fonctionnalité, si l'on utilise le classement préalable, mais entre des thèmes qu'il est plus facile de cataloguer sous les rubriques de "droite" et de "gauche", non au sens politicien du terme mais à un sens plus profond. A la droite peuvent être associés les thèmes de refus du laxisme dans la société ("simplifier les choses n'est pas une solution : on commence par l'orthographe mais après que voudrions-nous simplifier ?"), de critique de l'enseignement ("c'est tout l'enseignement qu'il faut revoir : les relations élèves professeurs qu'il faut repenser pour que le message passe") et de mise en avant de la notion de patrimoine intangible ("on ne doit pas révolutionner la langue française : il faut lui laisser un passé").

A la gauche peuvent être associés le souci de l'accès du plus grand nombre à la possibilité d'écrire ("rendre l'orthographe plus accessible, c'est faciliter l'accès de plus de personnes à la langue, à la lecture"), le souci de la simplification et l'insistance sur l'évolution nécessaire à la vie ("mieux vaudrait accepter le principe d'une évolution régulière plutôt que de procéder à de grands débats et révisions déchirantes à de rares moments de tension").

Cette opposition est mise en avant par le premier facteur et elle structure notre population, cependant il faut bien voir que cette population qui crée l'axe gauche/droite n'est pas la plus importante. En effet un retour aux données permet de faire les comptes suivants :

Effectif de ceux qui possèdent l'un des thèmes :

de droite : 36

---

<sup>22</sup> Ce tableau comprend 44 colonnes correspondant aux 44 modalités de réponse aux questions fermées et 17 lignes correspondant aux 17 thèmes : à l'intersection d'une ligne et d'une colonne se trouve le nombre de fois où un thème donné est associé à une réponse donnée. Il s'agit d'un "tableau thématique des questions", analogue au "tableau lexical des questions" (cf. Cibois 1989b). Quand le vocabulaire est suffisamment abondant, on peut rajouter celui-ci en lignes supplémentaires du tableau thématique des questions.

de gauche : 31

Comme il est possible d'avoir émis un thème de droite et un thème de gauche dans le même texte, il faut considérer les 4 cas logiquement possibles sur les 110 répondants ayant répondu à la question ouverte :

Thèmes de droite seuls : 20

Thèmes de gauche seuls : 25

Thèmes de gauche et de droite en même temps : 11

Aucun des 6 thèmes de droite ou de gauche : 54

On voit à partir de ces comptages que l'opposition gauche/droite est structurante tout en étant minoritaire. C'est un groupe de 45 individus sur 110 qui crée l'opposition de par la présence simultanée d'un petit nombre de thèmes.

On pourrait dire évidemment qu'il est bien difficile de tirer des conclusions d'un si petit nombre d'individus, à l'intérieur d'une enquête non représentative : on doit faire cependant remarquer que ces résultats sont cohérents avec l'enquête de la revue Lire où l'on nous dit que la volonté de réforme de l'orthographe est plus le fait de gens situés politiquement à gauche qu'à droite.

### Le débat "herméneutique"

Restent donc 11 thèmes : si certains sont plutôt du côté de la droite, ils sont soit de l'ordre de l'enracinement comme "étymologie", "charme" ; soit de l'ordre de la fonctionnalité comme "difficulté", "bizarreries", "réforme de la grammaire", soit de l'ordre de la controverse comme "cela forme la logique", "pas de décret arbitraire", "attention à la cohabitation de deux orthographes". Certains thèmes sont plutôt du côté de la gauche mais sont aussi de l'ordre de l'enracinement "importance de l'histoire", "beauté du mot" ; soit de l'ordre de la fonctionnalité : "des modifications sont à envisager".

Ceci manifeste que pour la majorité des répondants, il faut réussir à articuler ensemble l'enracinement et la fonctionnalité, c'est à dire résoudre cette quadrature du cercle qu'est pour toute culture le respect de ses racines culturelles sans que cela étouffe la nécessaire adaptation des outils culturels au présent.

Je pense que l'on peut légitimement nommer cette opposition "herméneutique" dans la mesure où toute évolution culturelle est une réinterprétation du passé, rendue nécessaire par les nécessités du présent. Il faut réutiliser le passé, l'interpréter pour rendre efficace le présent : il y a là un défi qui, s'il est plus visible au niveau de l'orthographe, se présente de manière analogue en ce qui concerne l'art, la religion et la culture en général. Si l'opposition herméneutique est moins visible que l'opposition gauche/droite, c'est qu'elle est à l'intérieur de chaque individu et que l'opposition politique n'en est que la partie rendue visible par les décisions pratiques à prendre face à un problème donné, décisions qui structurent le champ.

## II Place du latin dans l'enseignement

On a vu plus haut que le fait que le français utilise l'alphabet latin est la cause principale des difficultés de notre orthographe. Ce point n'est que l'une des traces de l'importance

historique du monde romain dans notre culture. Culture est pris ici comme synonyme de "culture occidentale", de cette démarche historique qui a deux sources, la culture grecque à qui l'on doit la science, la démocratie, la philosophie, la tragédie, la sculpture, l'architecture, etc., et la culture judéo-chrétienne à qui l'on doit la notion de sens de l'histoire.

Ces deux cultures ont été fondues dans la "voie romaine" (Brague 1992), c'est à dire dans cette synthèse spécifique qui nous a donné les langues romanes, l'église romaine, le droit romain, le saint empire romain et germanique, Byzance deuxième Rome, Moscou la troisième, les sœurs latines, l'Amérique latine, en un mot la latinité.

Que nous sentions proche de cette latinité, une des meilleurs preuves en est ce désir apparemment sans raison convaincante qu'ont les parents de niveau socialement élevé de faire apprendre le latin à leurs enfants, problème sur lequel je compte faire une réexploitation de l'enquête 1991 greffée au Panel DEP de l'Education nationale.

On entend dire sur les raisons de ce choix, en plus de motivations culturelles, qu'il s'agit d'une motivation distinctive ou d'une stratégie pour être dans les meilleures filières ou bien encore de l'apprentissage d'une "gymnastique de l'esprit". Je pense pouvoir montrer qu'il s'agit là de fausses motivations qui comblent un vide, un point obscur de notre vision de la culture.

En effet pour ce qui est de la motivation distinctive, elle peut tout aussi bien se faire dans des langues rares, le russe autrefois, le japonais aujourd'hui, qui ajoutent une motivation fonctionnelle ou exotique à la recherche de la distinction. Pour la recherche d'une filière d'excellence, le choix de l'allemand en première langue est infiniment plus efficace puisqu'il est obligatoire d'avoir une première langue vivante alors que faire l'option latin ne permet pas d'être dans une filière spécifique. Enfin pour ce qui est de la gymnastique de l'esprit, expression inventée par Jules Simon quand il a voulu faire abandonner le thème latin (et le discours latin) au profit de la version (Simon 1874), cette fonction de formation est faite de manière bien plus efficace aujourd'hui par les mathématiques ou par des pratiques ludiques comme les échecs ou la programmation informatique.

L'argumentation de l'intérêt pour le latin par l'accès direct aux textes ne tient pas : l'enseignement actuel du latin est orienté vers la version, non vers la lecture cursive des auteurs, pratique impensable pour des élèves en fin de cursus. De plus le travail de traduction est bien mieux fait par des experts qui diffusent tous les auteurs ayant écrit en latin et non le seul petit nombre des auteurs scolaires.

Mon hypothèse de départ dans ma réexploitation à venir est que la motivation à faire faire du latin est associée au fait de vouloir insister sur l'enracinement historique. On ne pourra pas le vérifier sur l'enquête liée au Panel cité, mais il est fort probable que le désir de faire faire du latin à ses enfants soit lié à ces deux indicateurs que sont le vote à droite et l'indicateur de pratique religieuse comme le laisse penser au moins le fait que l'étude du latin soit plus forte dans l'enseignement privé que dans l'enseignement public dont on sait qu'il est souvent confessionnel et s'adresse à un public plus souvent orienté politiquement à droite.

Quant à la motivation de l'étude du latin qui est que cette étude permettrait de mieux comprendre l'orthographe du français, il faut vraiment la prendre pour un symptôme. Elle signifie que l'on a l'intention d'étudier les racines du français et que l'on montre par-là que l'on reconnaît aux Anciens, dont nous dérivons, une supériorité sur les Modernes que nous sommes. Si en effet il n'y avait que pur calcul rationnel, il est évident que c'est l'apprentissage direct qui serait le plus efficace. Les "lois" de l'orthographe française sont dues à des interactions entre un système phonologique différent de celui du latin et un alphabet latin qui notait un grand nombre de racines latines. Les résultats ont formé un système nouveau dont

l'étude propre (l'orthographe du français) est ce qu'il y a de plus rentable pour l'apprentissage de l'orthographe.

En conclusion de toutes ces hypothèses, on peut dire que le désir de faire faire du latin aux enfants n'est raisonnable que dans une perspective de culture au sens de R. Brague (1992) : nous restons latins dans la mesure où nous continuons à recevoir notre héritage culturel d'autres que nous. Dans cette perspective il n'est pas possible de faire l'inventaire de ce que nous avons reçu, de traduire en langage d'aujourd'hui ce qui peut être encore utile et de jeter le reste comme on le fait quand on doit débarrasser une demeure ancienne. Pour filer cette métaphore, on doit plutôt considérer notre passé comme un grenier où l'on vient puiser, que l'on range, mais que l'on ne débarrasse pas. De ce point de vue, faire du latin aujourd'hui, c'est vouloir conserver un lien symbolique (et non efficace) avec notre passé culturel : il est possible que dans l'enseignement ce lien ne soit plus conservé que dans une filière spécialisée dont l'enseignement soit le seul débouché. Cette perspective vient en tout cas d'être proposée dans une revue d'enseignants (Mangin 1993).

### III Un courant pédagogique : l'Ecole Nouvelle

Toujours dans le domaine du débat sur la culture, mais en ce qui concerne sa transmission, j'ai été amené à m'intéresser à l'éducation par le biais de l'Education nouvelle telle que je l'ai vue pratiquée dans une Ecole Nouvelle "La Source" à Meudon.

La Source, Ecole "Nouvelle" : il ne faut pas faire d'erreur sur cette nouveauté de La Source qui, née après la dernière guerre, n'est plus une école récente. Le terme marque l'affiliation de l'Ecole à un courant pédagogique, celui de "l'éducation nouvelle". Ce terme apparaît pour la première fois à la fin du siècle dernier, mais la réalité, c'est à dire un courant pédagogique qui conteste l'école de masse en train de se créer, est contemporain de cette création.

Ce courant d'opposition dont les noms les plus connus sont Decroly ou Montessori va suivre, mais à l'envers, les hauts et les bas du système normal. Quand la société est contente d'elle-même, le courant pédagogique critique va en s'amenuisant, quand la société est désemparée, il se relève. On voit par exemple que les deux guerres mondiales, qui ont conduit notre civilisation au désespoir, lui ont fait perdre confiance dans son système éducatif. Ainsi, après la première guerre mondiale, beaucoup pensent que les adultes ont perdu tout droit moral à imposer leurs normes et qu'il faut repartir sur des bases nouvelles en isolant en quelque sorte les enfants de la perversion des adultes. C'est à cette époque que se sont spécialement développés en Europe le mouvement Scout et l'Action catholique des jeunes. Dans le domaine éducatif, c'est en 1920 que Roger Cousinet (l'inspirateur de La Source, cf. Raillon 1990) fonde une association "la Nouvelle Education". De même c'est après la seconde guerre mondiale que "L'Ecole nouvelle française", fondée en 1945 par Cousinet et Chatelain, reprend le flambeau, aiguillonnée par le désir de créer un nouveau monde sur des nouvelles bases : c'est ce qui conduira à la naissance de La Source, Ecole nouvelle, école qui se situe dans une lignée historique d'opposition au système standard d'enseignement.

#### La méthode Cousinet

Si la motivation à sortir du système scolaire standard réside dans le fait que celui-ci n'a pas su former des hommes qui aient évité le désastre des deux guerres, elle est purement négative et ne saurait suffire à fonder un nouveau type d'éducation. Pour cela Cousinet, entre les deux guerres, met au point une méthode destinée aux enfants du primaire dite "méthode de

travail libre par groupes" qui deviendra très vite la "méthode Cousinet". Cette méthode s'appuie sur les principes suivants :

1) contre une pédagogie traditionnelle où la passivité de l'élève est requise, où on lui demande d'être attentif, d'être obéissant, Cousinet oppose la manière dont se font les apprentissages chez l'enfant. L'enfant apprend à parler, à marcher, à jouer à des jeux compliqués, à se servir correctement d'instruments, non pas en suivant des cours, mais petit à petit, par imprégnation, par essais et erreurs, en agrandissant progressivement son rayon d'action.

"La vie de l'enfant a consisté, jusqu'à son entrée à l'école, en une activité continue, toujours avide d'être satisfaite, et qui se jette sur toutes les choses comme sur sa proie naturelle. Il entre à l'école, il se trouve en présence d'une personne nouvelle pour lui qui possède une quantité de choses également nouvelles (les diverses connaissances scolaires) : ces choses il les lui laisse entrevoir, et il lui interdit d'y toucher. Car le premier contact des enfants avec l'enseignement est une inhibition. A ces êtres qui ne vivent que pour agir, et qui ne peuvent vivre qu'à condition d'agir, l'action est interdite. L'activité d'abord sous sa forme matérielle. Il est interdit de bouger, de se lever, de changer de place, de parler, d'interroger, de regarder autre chose que ce qu'il est ordonné de regarder, d'entendre autre chose que ce qu'il est ordonné d'écouter, de toucher quoi que ce soit. L'activité sous sa forme spirituelle. Il est interdit d'interroger le maître (c'est le maître qui interroge), de solliciter une information sur un problème qui ne fait pas partie du programme ou à un autre moment que celui fixé par le programme. Il est même interdit aux enfants, si paradoxale que paraisse cette information, de s'instruire. Car ces choses nouvelles que le maître possède, que les enfants ont entrevues, dont ils désirent s'emparer à leur tour, le maître ne les montre qu'une à une, et comme à regret. Il dessine sur le tableau noir, ou il distribue des livres où sont dessinés ces choses mystérieuses qu'il appelle des lettres, mais il n'en montre qu'un peu à la fois, et avec une lenteur infinie. La science, le vrai butin des enfants, est présentée avec la même lenteur, d'une façon analytique, et dans un ordre rebutant."<sup>23</sup>

2) contre une pédagogie traditionnelle où l'enseignement se fait d'une manière strictement individuelle, il faut respecter le mode d'apprentissage naturel des enfants qui est collectif. "Dès le retour dans la classe, dès que les enfants se retrouvent en présence du maître et sous sa direction, l'activité redevient individuelle. Autrui est comme s'il n'existait pas. Il est interdit d'en prendre conscience. Chaque enfant doit agir seul, sans connaître le voisin, sans l'aider, sans être aidé de lui, sans lui parler, sans l'écouter. La discipline scolaire l'isole dans une cellule imperméable. (...) Devenu social, imprégné de socialisation, il ne peut pas redevenir un être individuel. Il essaie par tous les moyens de résister à cette autorité magistrale qui l'empêche de mener sa vie socialisée : il "souffle", il "copie", et surtout il parle, à telle enseigne qu'on pourrait dire, sans excès, que le maître passe presque moins de temps à parler lui-même qu'à empêcher ses élèves de parler."<sup>24</sup>

Cousinet fait donc se mettre au travail ses élèves par groupes librement choisis, sur des sujets librement choisis mais après avoir proposé un certain nombre de stimulations.

## Actualité de Cousinet

Les descriptions que Cousinet fait de l'enseignement standard nous semblent périmées : une pédagogie basée sur les centres d'intérêt, active, collective s'est effectivement substituée aux méthodes traditionnelles dans de nombreuses écoles primaires. A l'image de

---

<sup>23</sup> Cousinet, Une méthode de travail libre par groupes, Paris, Cerf, 1943, p.21-22.

<sup>24</sup> Idem p.44-45.

l'enseignement délivré comme dans une caserne, s'est substitué l'image d'une ruche bourdonnante où l'instituteur va de groupe en groupe pour répondre aux questions.

On peut donc dire que d'une certaine façon Cousinet a réussi puisque ses intuitions ont été reconnues : les "activités d'éveil" sont devenues parties prenantes des programmes officiels et le contact avec la nature, autre thème de l'Ecole nouvelle, a été repris en compte officiellement à travers les classes "externées", classes de neige d'abord, créées à Vanves par le docteur Fourestier dans les années cinquante puis largement étendues depuis.

Si d'une certaine façon, l'enseignement primaire français s'est suffisamment renouvelé et a en quelque sorte "digéré" le courant de l'Ecole nouvelle, la situation est beaucoup moins claire pour le collège et le lycée. Tout d'abord, il n'y a pas de "méthode Cousinet" pour ces âges : le groupe, qui est fondamental comme élément de socialisation entre 8 et 11 ans, n'a plus le même rôle ensuite. L'adolescent s'individualise et bien plus, réagit contre le groupe, même s'il en est parfois prisonnier. De la même façon, le deuxième principe de Cousinet relatif aux apprentissages "naturels" ne s'applique plus ici. Les apprentissages faits de la 6e à la terminale ne peuvent plus être assimilés à l'apprentissage par imprégnation de la langue maternelle. Il s'agit d'un ensemble de connaissances et de savoir-faire liés à des disciplines scientifiques ou littéraires : il devient difficilement pensable d'apprendre la physique par libre expérimentation où les mathématiques par essais et erreurs.

Confronté à ce problème, La Source a réagi de plusieurs façons que nous allons maintenant envisager.

## L'individu contre la société

La Source a subi de plein fouet l'impact de Mai 68. La critique de l'Ecole était partout : d'abord d'une manière globale comme instrument d'une société oppressante qui brime l'individu (Ivan Illich, *Une société sans école*, 1971) et, d'une manière plus spécifique comme instrument de reproduction d'une société capitaliste (Baudelot et Establet, *L'Ecole capitaliste en France*, 1971). De ce fait, La Source a été très marquée par le courant de la "pédagogie institutionnelle" (Vasquez et Oury, *Vers une pédagogie institutionnelle*, 1972) qui inversement reconnaît sa dette envers la pédagogie active du courant de l'Ecole Nouvelle.

Pour la pédagogie institutionnelle, les méthodes actives ont une efficacité parce qu'elles remettent en cause l'institution scolaire (et à travers elle l'institution sociale en général) : si l'on prend conscience par une "analyse" de caractère non naturel, "institué" dans le passé de l'institution scolaire, on se libère de la pression sociale et on donne à l'individu des possibilités d'épanouissement, de réussite personnelle. La pédagogie institutionnelle, dans la mouvance idéologique de l'époque, a voulu s'appuyer sur les désirs, les aspirations de l'individu, contre l'aspect répressif de l'institution scolaire (et sociale en général). A la limite, l'apprentissage d'une connaissance en vient à poser problème car il pouvait être considéré comme une violence faite par l'enseignant sur l'enseigné. C'est pourquoi, dans un livre récent<sup>25</sup>, cette orientation de La Source a été critiquée d'un point de vue analytique comme "l'illusion qui consiste à imaginer que le désir à pouvoir de tout"<sup>26</sup>.

Les enseignants actuels ont bien conscience qu'il est facile aujourd'hui, avec le recul du temps, de critiquer ce courant qui, puisqu'il n'a pas pu faire la révolution, changer la vie, a voulu cependant changer l'école et faire ainsi un travail souterrain qui ferait naître un avenir plus radieux. Ils ont bien conscience aujourd'hui que l'Ecole et la Société sont solidaires et que

---

<sup>25</sup> Florence Giust-Desprairies, *L'enfant rêvé*, Paris, Armand Colin, 1989.

<sup>26</sup> Idem, p.207.

préparer l'enfant à entrer dans la société, c'est lui donner des outils de connaissance mais c'est aussi l'intégrer aux rouages de la société par le biais de sa réussite scolaire sanctionnée socialement.

## La réussite individuelle mais conviviale

De ce fait, dans les années récentes, d'autres attentes se sont fait jour chez les parents : puisqu'il faut bien réussir son bac sous peine de marginalité sociale, autant le faire dans de bonnes conditions et l'on souhaite trouver à La Source un enseignement traditionnel, efficace, mais où les rapports sociaux soient autres. La relation pédagogique est personnalisée, quasi-familiale, les problèmes sont débattus et résolus en douceur. En matière de sanction on a toujours une dernière chance et ce principe est d'application récursive (d'où l'existence de rattrapage du rattrapage du...etc.). C'est à dire que l'institution, par sa taille limitée, par son souci de convivialité cherche à gommer les arêtes d'une société trop dure. De même que la famille est dans notre société l'instance ultime de protection, d'aide, de soutien, de même dans le domaine scolaire ce rôle est dévolu à La Source. Certains jugent d'ailleurs que cette attente de réussite confortable est une synthèse de l'idéal soixante-huitard d'épanouissement de l'individu et du désir contemporain de réussite dans la compétition sociale, que c'est en quelque sorte "la réussite à des prix-Libé".

Certains des enseignants se demandent si ce n'est pas une fausse piste, car c'est demander à une école de résoudre individuellement un problème collectif, c'est demander à des enseignants de trouver la solution de la contradiction entre des programmes exigeants, des élèves rétifs et des familles désireuses de réussite : petit à petit la convivialité disparaîtra sous la pression scolaire et seule restera la poursuite de la compétition. La Source deviendra alors une bonne école privée pour enfants de cadres supérieurs garantissant un bon taux de succès. Cet objectif est jugé en soi honorable mais il devra se payer de la disparition tendancielle de la convivialité : on ne peut vouloir l'intégration dans le système sans en payer le prix à terme.

## Enquête en cours

Voici donc en quelques mots la situation d'une école représentative du courant de l'Education nouvelle. L'enquête faite actuellement avec F. de Singly dans le cadre d'un enseignement de Deug teste plusieurs hypothèses :

### 1) principe de cohérence éducative

Puisque le choix de mettre un enfant dans une école nouvelle est le signe de l'accent mis par les parents sur le respect de l'autonomie de l'enfant, il faut tester s'il y a cohérence des choix éducatifs et voir si dans les pratiques éducatives elles-mêmes, on retrouve des choix en faveur de l'autonomie de l'enfant. A cette fin, non seulement on a posé un assez grand nombre de questions, mais on a également interrogé à des fins comparatives, une population socialement équivalente (classes moyennes et supérieures salariées) ayant mis ses enfants dans des collèges publics où ces classes sociales sont plutôt dominantes. Comme l'école nouvelle doit rendre apte les élèves à obtenir les mêmes diplômes que ceux de l'enseignement traditionnel, on fait l'hypothèse que l'on rencontrera un continuum de réponses (et de pratiques) qui iront des plus tournées vers le développement de l'enfant, jusqu'aux plus traditionnelles.

### 2) modification des choix éducatifs

Le modèle éducatif traditionnel met l'accent sur la scolarité et sur l'acquisition d'un haut diplôme qui ouvre la voie aux filières les plus prestigieuses. On peut faire l'hypothèse qu'à travers l'insistance sur l'autonomie, on assiste à la substitution du modèle précédent par un modèle mettant l'accent sur l'acquisition de capacités relationnelles qui sont jugées

indispensables aujourd'hui dans la société bien plus que des acquis scolaires. On peut faire l'hypothèse cependant que ce nouveau modèle n'est pour beaucoup de parents qu'une substitution imposée par une élimination non acceptée des filières nobles du système scolaire. On peut cependant faire aussi l'hypothèse que ce modèle mettant l'accent sur les capacités relationnelles, est en train de se substituer globalement au précédent et que l'Education nouvelle est le lieu d'une innovation qui va se répandre ailleurs comme cela a été le cas dans le primaire ce qui rendrait les deux hypothèses précédentes non contradictoires.

3) une stratégie différente selon le sexe

On peut également faire l'hypothèse que ce sont plutôt aux filles qui sont proposées les nouvelles manières de se construire une identité sociale en mettant l'accent sur les capacités relationnelles au détriment de l'ancien modèle féminin mettant l'accent sur les capacités morales (Cf. F. de Singly 1992).

4) un choix de parents non homogènes

Enfin on peut faire l'hypothèse que l'Ecole nouvelle, si l'on y trouve une tension entre le désir d'autonomie pour l'enfant et un désir de réussite scolaire, c'est parce qu'y viennent d'une manière préférentielle des parents non homogènes en ce qui concerne le milieu professionnel. Que par exemple le père insiste plutôt sur l'aspect scolaire parce que lui même est ingénieur et passé par le moule des grandes écoles et la mère est dans un milieu professionnel qui insiste sur les valeurs d'épanouissement et d'autonomie comme le milieu universitaire ou les métiers ayant rapport avec la psychologie.

L'enquête a permis de recueillir environ 500 réponses dont une centaine provenant d'écoles nouvelles. Son dépouillement est en cours.

## IV Prospective

Ces recherches sur des thèmes liés à la culture ou à sa transmission me permettent aujourd'hui de présenter quelques hypothèses de travail qui pourront me guider dans les années à venir :

1) des problèmes sociaux différents comme celui de la Culture, des rapports avec les autres cultures, de l'enseignement, de l'éducation, de la langue, de l'art, de la religion sont sous-tendus par un même débat herméneutique ;

2) ce débat porte sur "réception" pour aujourd'hui des sources anciennes de notre culture, sur les modifications nécessaires que les besoins du temps présent entraînent en ce qui concerne cet héritage ;

3) si dans chaque domaine les acteurs sociaux s'organisent en camps opposés qui mettent l'accent soit sur l'adaptation, soit sur la conservation, cette opposition se retrouve d'abord à l'intérieur de chaque individu. Tout le monde souhaite maintenir les deux aspects : comme chacun met le point d'équilibre à un niveau différent, le regroupement de ceux qui font des choix plus adaptatifs ou plus conservateurs crée des groupes "novateurs" ou "conservateurs" ;

4) le champ politique semble partagé par le même débat mais sa spécificité fait que l'opposition gauche/droite, si elle a des points communs avec le débat herméneutique, ne lui est pas superposable et demande une analyse spécifique.

\*\*\*

## Nota bene

Ce texte a été orthographié en suivant les recommandations du Conseil supérieur de la langue française approuvées à l'unanimité par l'Académie française le 3 mai 1990 et que celle-ci a souhaité, le 17 janvier 1991 soumettre "à l'épreuve du temps". Ces modifications portent surtout sur des simplifications d'accent (absence d'accent circonflexe sur le "i" et sur le "u") et sur quelques régularisations ("évènement" comme "avènement").

Ces modifications voulues ne préjugent pas de quelques modifications involontaires, non voulues ni par le Conseil supérieur de la langue française, ni par l'Académie française.

\*\*\*

## Références bibliographiques

BARDET, J.-P.

1983 *Rouen aux XVIIe et XVIIIe siècles*, SEDES, Paris.

BENZECRI, J.-P.

1973 *L'analyse des données*, Dunod, Paris.

BERNARD, J.-M.

1986 "Méthodes d'inférence bayésienne sur des fréquences", *Informatique et Sciences Humaines*, n°68-69, p.89-133.

BERTIN, J.

1967 *Sémiologie graphique*, Mouton, Paris et La Haye.

1977 *La graphique et le traitement graphique de l'information*, Flammarion, Paris.

BOUDON, R.

1973 *L'inégalité des chances*, Armand Colin, Paris

BOUDON, R., LAZARSELD, P.

1965 *Le vocabulaire des sciences sociales*, Mouton, Paris et La Haye.

1966 *L'analyse empirique de la causalité*, Mouton, Paris et La Haye.

BRAGUE, R.

1992 *Europe : la voie romaine*, Critérim, Paris.

CIBOIS, Ph.

1971 "Analyse factorielle des correspondances appliquée à une enquête d'opinion", in *Communications présentées au colloque de l'Arbresle sur l'analyse des données, 21-22 avril 1971*, Musée National des Arts et Traditions Populaires - Centre d'Ethnologie Française, p.18-25.

- 1973 *Enquête sur la méthodologie et l'utilisation des moyens informatiques dans les recherches économiques et sociales*, Rapport de fin d'études ADISH/CORDES, convention de recherche 34/1972.
- 1975 "Le colloque de Rome sur les typologies d'amphores : d'une vision synthétique à un discours analytique", *Informatique et Sciences Humaines*, n°25, p.11-30.
- 1976 *Demande individuelle d'éducation. Etude de cas : France*, rapport OCDE, Paris.
- 1980 *La représentation factorielle des tableaux croisés et des données d'enquête : étude de méthodologie sociologique*, Thèse de doctorat de 3e cycle, Université Paris V.
- 1981 "Analyse des données et sociologie", *L'Année sociologique*, 31, p.333-348.
- 1982a "Tri-deux : une méthode post-factorielle de dépouillement d'enquête", *L'Année sociologique*, 32, p.61-80.
- 1982b "Le Natalisme national", *Esprit* (10), p.76-96.
- 1983a *L'analyse factorielle*, Presses Universitaires de France, Paris.
- 1983b "Méthodes post-factorielles pour le dépouillement d'enquête", *Bulletin de méthodologie sociologique*, n°1, p.41-78.
- 1984 *L'analyse des données en sociologie*, Presses Universitaires de France, Paris.
- 1985 "Belle-Ile : débat de presse sur un été chaud. Etude par la méthode synoptique", *Déviance et Société* 9 (4), Genève, p.313-332.
- 1986a "Introduction à la méthode Tri-deux", *Informatique et Sciences Humaines*, n°70-71, p.5-13.
- 1986b "Analyse secondaire d'une enquête sur la natalité", *Informatique et Sciences Humaines*, n°70-71, p.15-30.
- 1989a "Pour une science sociale synchronique", *La revue du MAUSS* n°4, p.70-84.
- 1989b "Eclairer le vocabulaire des questions ouvertes par les questions fermées : le tableau lexical des questions", *Bulletin de méthodologie sociologique*, n°26, p.12-23.
- 1989c "80%, est-ce possible ?", *Le Monde* 22 juin.
- 1989d "Orthographe et société" in J.Leconte et Ph.Cibois, *Que vive l'orthographe*, Seuil, Paris, p.67-130.
- 1991 "Enquêtes d'opinion sur l'orthographe", *Mots* n°28, p.86-98.
- 1992a *Soixante styles de vie et de pratiques culturelles*, rapport CERSOF/Université Paris V.
- 1992c "Mille accents mis ou omis", *Liaisons-HESO* n°19-20, p.29-34.
- 1992d "L'Ined, Le Bras et Calot", *Esprit* (1), p.159-165.  
à paraître
- 1993a "L'espace du chercheur : le graphique", *Communication aux journées d'études Ethologie-Sciences sociales* du 25 mars 1992, L'Harmattan.
- 1993b "Repenser les enquêtes par sondage : un modèle structuré rend possible de nouvelles méthodes de dépouillement", (en anglais) in K. Van Meter (Ed.), *Interrelations between Methods of Analysis and Interpretation of Results*, Sage Publications.
- CIBOIS et al.
- 1972 (avec Jean Leclant et Michel de Virville), "Le traitement sur ordinateur des inscriptions méroïtiques", *Informatique et Sciences Humaines*, n°15, p.5-18.
- 1976 (avec Janina Markiewicz-Lagneau), *Les étudiants dans l'enseignement supérieur court*, OCDE, Paris, également publié en anglais sous le titre *Students in short-cycle higher education*.

- 1988a (avec Jean-Jacques Droesbeke), "La croissance du nombre des bacheliers est-elle modélisable et prévisible", *Revue française de sociologie*, 29 (3), p.425-445.
- 1988b (avec Jean-Marc Bernard), "Introduction à l'utilisation du logiciel d'inférence bayésienne sur les fréquences IBF2XK", *Bulletin de méthodologie sociologique*, N°20, p.12-24.
- CLERO J.-P.,  
1988 "Un instrument de mesure des croyances. La règle de Bayes", *Histoire & Mesure*, 3 (4), p.491-513.
- COLLIOT-THELENE, C.  
1990 *Max Weber et l'histoire*, Presses Universitaires de France, Paris.
- DAVAL, R, et al.  
1967 *Traité de psychologie sociale*, Tome I, Presses Universitaires de France, Paris.
- DESROCHE, H.  
1971 *Apprentissage en sciences sociales et éducation permanente*, Les éditions ouvrières, Paris.
- DOISE, W., CLEMENCE, A., LORENZI-CIOLDI, F.  
1992 *Représentations sociales et analyse des données*, Presses Universitaires de Grenoble, Grenoble.
- DONNAT, O. et COGNEAU, D.  
1990 *Les pratiques culturelles des Français, 1973-1989*, La Découverte/La Documentation Française, Paris
- DROUARD, A.  
1992 *Une inconnue des sciences sociales, la Fondation Alexis-Carel : 1941-1945*, Institut National d'Etudes Démographiques - Maison des Sciences de l'Homme, Paris.
- FEYERABEND, P.  
1979 *Contre la méthode*, Seuil, Paris.
- GARDIN, J.-Cl.  
1991 *Le calcul et la raison*, EHESS, Paris.
- GRIZE, J.-Bl.  
1990 *Logique et langage*, Ophrys, Gap.
- GUILBAUD, G.,Th.  
1985 *Leçons d'à-peu-près*, Ch.Bourgeois, Paris.
- HERPIN, N.  
1986 "Socio-styles", *Revue française de sociologie*, 27, p.265-272.
- LAGNEAU J., NETTER, M. et LORIEUX J.

- 1973 *Les étudiants des Instituts universitaires de technologie en France, rapport OCDE.*
- LEBART, L., SALEM, A.  
1988 *Analyse statistique des données textuelles*, Dunod, Paris.
- LEBRAS H.  
1991 *Marianne et les lapins*, Olivier Orban, Paris.
- LECOUTRE, B.  
1984 *L'analyse bayésienne des comparaisons*, Presses Universitaires de Lille, Lille.
- LEVI-STRAUSS, Cl.  
1952 "Le Père Noël supplicié", *Les Temps Modernes*, n°77, p.1572-1590.
- Lire  
1989 "Sondage - Faut-il réformer l'orthographe", *Lire*, n°162, p.47-61.
- MANGIN, D.  
1993 "Jean Bollack, l'Oedipe roi de Sophocle, Paul Veyne, la Société romaine", *L'Ecole des lettres II*, 1992-1993 (8), p.93-98.
- MEAR, P. et MERLE, P.  
1991 "Problèmes de modélisation prévisionnelle : l'exemple de la croissance du taux de bacheliers", *Revue française de sociologie*, 32 (2).
- Ministère de la culture  
1990 *Nouvelle enquête sur les pratiques culturelles des français en 1990*, La Documentation Française, Paris.
- RAILLON, L.  
1990 *Roger Cousinet, une pédagogie de la liberté*, Armand Colin, Paris.
- REINERT, M.  
1987 "Classification descendante hiérarchique et analyse lexicale par contexte. Application au corpus des poésies d'A.Rimbaud", *Bulletin de méthodologie sociologique*, n°13, p.53-90.
- ROUANET, H., LE ROUX, B. et BERT, M.-Cl.  
1987 *Statistique en sciences humaines : procédures naturelles*, Dunod, Paris.
- REUCHLIN, M.  
1964 *Méthodes d'analyse factorielle à l'usage des psychologues*, Presses Universitaires de France, Paris.
- REUCHLIN, M. et BACHER, F.  
1969 *L'orientation à la fin du premier cycle secondaire*, Presses Universitaires de France, Paris.
- RICOEUR, P.

1983 *Temps et récit*, Tome I, Seuil, Paris.

ROUDINESCO, E.

1986 *Histoire de la psychanalyse en France*, TI, 1885-1939, TII, 1925-1985, Seuil, Paris.

SIMON, J.

1874 *La réforme de l'enseignement secondaire*, Paris.

SINGLY de, F.

1992 "Les habits neufs de la domination masculine", *Esprit*.

SINGLY de, F.

1993 "Les jeunes et la lecture", *Les dossiers Education et Formations*, n°24, Ministère de l'Education et de la Culture.

\*\*\*

## Postface de 2010

Au cours de la soutenance, Roger Establet m'avait demandé ce que je pensais de la régression logistique : je n'en pensais rien du tout, ne m'étant pas intéressé jusqu'à présent à la méthode mais j'étais un petit peu vexé d'avoir du détourner la question en prétextant de la complexité du problème. J'ai donc étudié à fond la question et j'ai découvert qu'il était possible de repérer des résultats très proches de ceux donnés par les méthodes classiques simplement en construisant des tableaux croisés et en prenant des moyennes d'écart bruts à l'indépendance. C'est ce que j'ai appelé ensuite la « méthode tabulaire » qui a été présentée et défendue dans les deux articles suivants :

- 1999 "Modèle linéaire contre modèle logistique en régression sur données qualitatives", *Bulletin de méthodologie sociologique*, n°64, p.5-24.

- 2000 "Observation et modèle linéaire ou logistique : réponse à Aris et Hagenaars", *Bulletin de méthodologie sociologique*, n°67, p.54-64.

L'avantage de cette méthode est que du point de vue pédagogique elle permettait à des sociologues familiers des tableaux croisés de comprendre le principe de la régression « toutes choses égales par ailleurs » : on voyait bien qu'il s'agissait, comme pour l'analyse factorielle d'une approximation des données. Cependant il fallait faire comprendre ce que c'était qu'un Odds ratio, quelle était la logique de cet indicateur et comment l'utiliser convenablement.

Comme les PUF souhaitaient que je fasse une refonte complète de mon *Que sais-je ?* sur l'analyse factorielle, j'ai tiré profit de tous les enseignements donnés sur cette question, de l'évolution des mœurs en ce qui concerne l'utilisation des méthodes informatisées et j'ai d'une part intégré les méthodes de régression logistique dans Trideux (version 3) et refait un nouveau *Que sais-je ?* qui portait à la fois sur l'analyse factorielle et sur la régression logistique :

- 2007 *Les méthodes d'analyse d'enquêtes*, Presses Universitaires de France, Paris, coll. "Que sais-je ?", n°3782.

Par rapport au premier ouvrage, le second explore davantage les bonnes pratiques que doivent avoir les chercheurs, en particulier en analyse factorielle pour en déjouer les pièges que j'avais rassemblés dans :

- 1997 "Les pièges de l'analyse des correspondances", *Histoire & Mesure*, 12 (3/4), pp. 299-320.

L'idée qu'un plan factoriel représente un type-idéal wébérien a été réutilisé ensuite par Bernard Lahire (*La culture des individus*, La découverte 2004, p. 132-136) pour expliquer comment les résultats de Bourdieu étaient pour une part dus à un effet de ce genre et comment on pouvait au contraire, par des dénombrements des pratiques, voir des dissonances culturelles parmi les individus.

En parallèle à ces travaux sur les méthodes, j'ai continué à travailler sur les pratiques culturelles dans deux directions :

- les publics du théâtre : à Amiens, les rapports entre l'université où j'étais et la Maison de la culture étaient très bons et la direction m'avait demandé (ainsi qu'à Jean Copans) d'essayer d'explorer la question d'une éventuelle désaffection des abonnés. L'examen du fichier des abonnés a permis de se rendre compte que ce n'était pas le cas et les contacts pris ont permis de passer un contrat avec la DEP du Ministère de la culture et de faire des entretiens approfondis d'une trentaine d'abonnés. Les résultats, publiés dans les textes ci-dessous ont permis d'élaborer une typologie qui oppose les adeptes du théâtre total (pour qui

les valeurs qu'ils trouvent dans la représentation théâtrale sont un écho de leurs préoccupations profondes politiques et morales) et les gens du théâtre (étudiants en voie de professionnalisation) et ceux pour qui le théâtre est un divertissement culturel assumé.

- 2003 "Les abonnés du théâtre : un public hétérogène", in Olivier Donnat (Dir.), *Regards croisés sur les pratiques culturelles*, La documentation française, p.171-187.

- 2003 "Comprendre les publics du théâtre : l'exemple des abonnés d'une scène nationale", in Olivier Donnat, Paul Tolila (Dir.), *Le(s) public(s) de la culture*, Presses de Sciences-Po, vol. 2, p.169-174.

- 2008 "Les publics du théâtre contemporain", *Communications*, n°83, p. 37-46.

- la question du latin

*Rosa, rosa, rosam, rosae, rosae, rosa...*La déclinaison latine berce encore les souvenirs de beaucoup, que ce soit sous sa forme d'origine ou dans la version chantée de Jacques Brel. Elle évoque la nostalgie d'un âge révolu, d'une époque où l'on apprenait des choses certes inutiles, mais étudiées sérieusement, où les lycées faisaient entrer dès la 6e dans le monde exotique du *de viris illustribus* en compagnie de Romulus "premier roi des romains" ce qui s'écrivait, à l'envers, "romanorum rex primus". Pour un enfant de 11 ans c'est un peuple qui parle une langue bizarre, qui ne sait pas traduire l'article et qui ne sait pas parler normalement puisqu'il met curieusement son verbe à la fin. Alors que les gens normaux disent que "Procas, roi des Albains eut deux fils, Numitor et Amulius", le début du *de viris* semble mettre tous les mots dans un panier et les sortir au hasard, laissant toutefois systématiquement le verbe à la fin, par pure provocation : "Proca, rex Albanorum, duos filios, Numitorem et Amulium, habuit."

Si j'évoque ces souvenirs d'enfance, c'est qu'à l'origine de cette recherche se posait une question qui était de savoir pourquoi l'on fait du latin dans l'enseignement secondaire. Quand j'ai moi-même fait du latin tout au long du secondaire, la question ne se posait pas, j'ai été inscrit d'autorité dans une section classique mais je suis à cette époque personnellement arrivé très vite à la conclusion que le latin ne servait à rien et que s'il y avait une œuvre de salubrité publique à assurer pour les générations futures, c'était la promotion de la suppression du latin dans le secondaire. N'ayant plus eu à me servir du latin, y compris à l'Eglise à la messe après l'époque du concile Vatican II, j'avais laissé la question en suspens jusqu'au jour où, à leur entrée en 4e, j'ai dû choisir de faire faire ou non du latin à mes enfants et c'est à ce moment que la question s'est posée à moi comme parent, mais aussi comme sociologue. Le choix du latin est assez fréquent : pour quelles raisons ? Ces raisons pouvant être avouées ou inavouées, conscientes ou non. Si, comme je l'avais cru au lycée, le latin ne servait à rien, pourquoi se maintenait-il avec tant de vigueur puisqu'après une génération, il était toujours là à être proposé à tous. Il y avait là une question qui me concernait directement, sans toutefois m'impliquer profondément : c'était une belle question.

En était-ce véritablement une ? Pas du tout car pour la plupart des gens que j'interrogeais la question ne se posait pas : c'était évidemment une aide pour la compréhension du français et une gymnastique de l'esprit. Ce n'était pas une bonne question non plus pour la plupart de mes collègues qui me regardaient avec un léger sourire pour me dire qu'enfin, on savait bien pourquoi les gens font faire du latin à leurs enfants, que c'était pour être avec les meilleurs et par souci de distinction sociale. Je n'ignorais évidemment ni les raisons avouées ni les raisons inavouables mais le travail du sociologue n'est pas de prendre au mot ce qui se dit sur une question, même si c'est un langage savant qui est utilisé. Voilà pourquoi je me suis intéressé à cette question : travail de routine mais qui, comme souvent dans la recherche, m'a conduit sur un tout autre terrain que celui que j'avais envisagé au départ.

J'ai rassemblé sur cette question beaucoup d'enquêtes, d'exploitation statistiques, de comparaisons internationales, d'études historiques mais, comme il me semblait qu'il fallait réformer ce qui se faisait, ce mélange de performatif et d'informatif a choqué les directeurs de plusieurs maisons d'éditions, même quand les directeurs scientifiques de collections avaient donné un avis favorable. Lassé de cette procédure, je diffuse aujourd'hui mes résultats sous la forme d'un carnet de recherche en ligne, [La Question du latin](#), qui a l'avantage de régulièrement me permettre d'informer les professeurs de latin, de discuter avec eux dans des colloques spécialisés, des émissions de radio.

- retour au point de départ

Lors d'un des derniers cours que je faisais avant de partir à la retraite, je discutais avec une étudiante de première année à une fin de cours quand brutalement elle me posa la question « c'est vrai que vous avez été prêtre ? ». Je suppose que pour elle, une écoute attentive de la part d'un enseignant devait supposer un don de soi surnaturel, mais je suis resté frappé par le fait qu'un évènement vécu par moi il y a maintenant 40 ans soit connu en première année. La mémoire collective est forte sur ces questions car pour mon compte, je ne me suis jamais vanté de cela (ni caché non plus d'ailleurs, cf. mon [article d'Esprit](#) de 1977 « Pourquoi on quitte l'Eglise »).

A l'occasion d'un séminaire commun du laboratoire Printemps de l'UVSQ animé par Sandrine Nicourd et portant sur les socialisation de l'engagement militant, j'ai essayé de montrer en m'appuyant en partie sur mon expérience personnelle que la vocation sacerdotale était une construction sociale qui était de ce fait un mélange de contrainte et de liberté. On verra comment ces deux aspects sont conjoints dans un chapitre du livre sur *Le travail militant* (PUR, 2009) « [La construction sociale de la vocation sacerdotale](#) »

Dans mon texte de 1993, dans la « prospective », il me semblait que la réception du passé, en particulier du passé religieux, passait par un *débat herméneutique* qui était clivant. Cela ne me paraît plus aussi sûr dans la mesure où à cette époque, le débat politique était encore effectivement clivé entre une gauche attaché aux Lumières et une droite attachée à la tradition en général. Depuis, l'idéal universaliste des Lumières est apparu pour beaucoup comme l'imposition d'une manière de voir occidentale (Le livre récent de Jacques Goody, *Le vol de l'histoire* en est un symptôme) et les idées de Gadamer s'imposent aujourd'hui comme quoi « le dépassement de tous les préjugés, cette exigence globale de l'*Aufklärung*, s'avérera être lui-même un préjugé, dont seule la révision frayera la voie à une compréhension appropriée de la finitude qui domine non seulement notre être, mais également notre conscience historique. »<sup>27</sup>

De ce fait, l'enseignement du religieux s'impose aujourd'hui : comment serait-il possible de comprendre le moindre débat politique en France sans connaître le débat religieux antérieur ; comment pourrait-on voir une cathédrale, un tableau à thème religieux, un texte ancien sans connaissance de la religion chrétienne ? Evidemment, cet enseignement doit être fait dans une perspective agnostique et respectueuse de la laïcité de l'enseignement. Ayant une bonne connaissance de ce corpus religieux, en même temps que la réforme que je juge nécessaire de l'enseignement du latin, je pousse à ce que les textes latins étudiés soient ceux de la Vulgate, des Pères de l'Eglise, du Moyen Age. On trouvera une tentative récente en ce sens dans la réponse que j'ai apporté à cette question « [Que dire de Noël dans l'enseignement ?](#) »

---

<sup>27</sup> *Vérité et méthode*, Seuil, 1996, p.297

# Table des matières

Introduction .....	1
Chapitre 1 : les éblouissements .....	1
Le musée des Arts et traditions populaires .....	1
La Maison des Sciences de l'Homme .....	2
Une enquête sur le Patrimoine .....	3
L'analyse empirique en sociologie .....	4
Etude d'une cohorte d'élèves .....	5
Le retour aux données .....	5
Sémiologie graphique .....	6
Premiers travaux informatiques .....	7
Informatique et Sciences Humaines .....	8
L'approximation .....	9
Chapitre 2 : sociologie de l'éducation .....	11
L'enseignement supérieur court .....	11
La représentation en surface des tableaux croisés .....	11
La demande d'éducation.....	14
80% de bacheliers .....	14
Analyses récentes.....	15
Comparaison PEM taux de liaison.....	15
Exemple d'utilisation.....	16
Conclusion .....	20
Chapitre 3 : analyse des données.....	21
L'expérience du LISH .....	21
Thèse de troisième cycle .....	21
L'analyse factorielle .....	23
L'analyse des données en sociologie.....	23
Le tri-factoriel des questions.....	25
Chapitre 4 : le logiciel Tri-deux .....	27
Tri-deux sous IBM 370.....	28
Tri-deux en micro-informatique : première version .....	28
Environnement d'assistance et de formation.....	29
Avatars .....	30
Tri-deux deuxième version .....	32
Chapitre 5 : analyse de contenu.....	34
Sources anciennes : le royaume de Méroë.....	34
Analyse de contenu d'un débat : Belle-Ile, été 1984.....	34

La méthode synoptique .....	36
Codage .....	36
Analyse et interprétation .....	37
Le logiciel Tri-deux Mots .....	38
Formalisations ultérieures .....	39
Chapitre 6 : démographie .....	39
Généalogie de l'INED .....	40
Une enquête de la Sofres sur la natalité .....	41
Chapitre 7 : inférence .....	42
Doit-on craindre les petits effectifs ? .....	43
Les techniques d'inférence bayésiennes .....	44
Chapitre 8 : science sociale synchronique.....	46
L'article de la revue du MAUSS .....	46
Les pratiques culturelles des français .....	47
Styles de vie .....	47
Aller plus loin ?.....	50
Chapitre 9 : sociologie de la culture.....	51
I La réforme de l'orthographe .....	51
Le paradoxe enracinement/fonctionnalité .....	52
Situation du débat .....	57
Le débat "herméneutique" .....	58
II Place du latin dans l'enseignement .....	58
III Un courant pédagogique, l'Ecole Nouvelle.....	60
La méthode Cousinet.....	60
Actualité de Cousinet .....	61
L'individu contre la société .....	62
La réussite individuelle mais conviviale .....	63
Enquête en cours .....	63
IV Prospective.....	64
Nota Bene .....	65
Références bibliographiques .....	65
Postface .....	70
Table des matières .....	73