Réseau A.D.O.C.

Amélioration Des Outils de Connaissance
pour la gestion urbaine dans les pays en développement

Document de travail n° 4

ANALYSE STATISTIQUE D'UN SYSTEME D'ECHELLES

Marie PIRON, ORSTOM - PARIS VI

Avril 1992

ORSTOM Institut français de recherche scientifique pour le développement en coopération

Le travail présenté dans la présente publication a fait l'objet d'une Thèse de Doctorat, sous la direction de Monsieur Jean-Paul BENZECRI, soutenue le 7 janvier 1991 à l'Université Paris VI.

PREFACE

En France comme dans de nombreux autres pays, les années quatre-vingt ont été marquées par le développement des systèmes d'information urbains. Leur mise en place et leur utilisation soulève un certain nombre de questions d'ordre diverses : choix, mobilisation et actualisation des informations, gestion, traitement et analyse de celles-ci, mais aussi problèmes humains, organisationnels et politiques soulevés par la mise en place de tels outils.

Dans cet ensemble de problèmes, ceux posés par la manipulation conjointe de données de natures, de qualités et d'échelles différentes tiennent une place particulière, de plus en plus importante avec la multiplication des sources d'information. A ces difficultés, se surimpose souvent, et tout particulièrement dans les villes des pays en développement, celles liées au caractère incomplet des données. Avec les nouveaux systèmes disponibles sur le marché, la résolution des problèmes informatiques posés par l'intégration et la manipulation d'informations diverses, correspondant à des niveaux différents, devient de plus en plus effective. Mais il importe que la réflexion et le développement des méthodes d'analyse aillent de pair avec le progrès informatique.

En effet, la multiplicité des informations intégrées dans les systèmes d'information urbain posent avec une acuité particulière la question, classique, des niveaux de perception et d'analyse des systèmes géographiques. L'analyse de tout système géographique, qu'il soit urbain ou non, repose sur des informations relatives à des agrégats, entités spatiales ayant une réalité sociale, administrative, voire économique. Or l'agrégat n'est qu'une vision partielle : le passage, les changements de pertinence, les accentuations ou les pertes d'information liés à la hiérarchie spatiale introduite dans les analyses urbaines restent des points importants à étudier, tout aussi fondamentaux que l'introduction du temps dans les analyses urbaines.

Les méthodes statistiques que propose Marie PIRON dans cet ouvrage ouvrent des perspectives intéressantes de compréhension et d'analyse qui devraient réjouir tous ceux qui sont confrontés à ces questions récurrentes dans l'analyse des systèmes urbains : "Comment croiser des informations définies sur des niveaux différents ?", "Que se passe-t-il lorsque je change de niveau d'analyse ?". Aussi, malgré l'aspect aride que peut recouvrir pour certains les quelques développements mathématiques présents dans cet ouvrage, sa lecture ne peut qu'être vivement conseillée : les travaux de Marie PIRON constituent un apport appréciable pour tirer parti de la diversité de l'information disponible sur les villes.

Françoise DUREAU
et Christiane WEBER
Avril 1992

SOMMAIRE

INTRODUCTION GENERALE 5

PREMIERE PARTIE : **EXEMPLE D'ELABORATION D'UNE STRUCTURE HIERARCHIQUE DEL'INFORMATION** 13

Introduction 15
Chapitre 1.1: Stratégie d'enquête 17
Chapitre 1.2: Préparation de la collecte et de l'analyse des données 21
Conclusion : Vers l'exploitation de la structure hiérarchique de l'information 29

DEUXIEME PARTIE : **CONCEPTION DU TABLEAU POUR EXPLOITER UNE STRUCTURE HIERARCHIQUE DE L'INFORMATION** 31

Introduction 33
Chapitre 2.1: Des données aux tableaux 35
Chapitre 2.2: Principe de construction du tableau d'information 41
Chapitre 2.3: Les divers tableaux du processus d'agrégation 55
Conclusion : Vers l'analyse d'un système d'échelles 65

TROISIEME PARTIE : **UNE METHODE POUR L'ANALYSE D'UN SYSTEME D'ECHELLES** 67

Introduction 69
Chapitre 3.1: Les principes de la décomposition de l'inertie 71
Chapitre 3.2: Analyse sur les éléments de la décomposition de l'inertie 79
Chapitre 3.3: Aides a l'interprétation du système d'échelles 89
Conclusion : Proposition d'une méthode d'analyse d'un Système d'échelles 99

QUATRIEME PARTIE : **UNE APPLICATION DE LA METHODE D'ANALYSE DU SYSTEME D'ECHELLES** 101

Introduction 103
Chapitre 4.1: Contexte expérimental : enquête sur Ouagadougou 105
Chapitre 4.2: L'analyse du système d'échelles 111
Chapitre 4.3: Influence du système d'échelles 131
Conclusion 137

CONCLUSION GENERALE 141

ANNEXE 1 : **TECHNIQUES D'ENQUETE** 147

ANNEXE 2 : **COMPLEMENTS STATISTIQUES POUR L'ANALYSE DU SYSTEME D'ECHELLES** 183

BIBLIOGRAPHIE 203

TABLE DES MATIERES 209

Introduction générale

LES SYSTEMES D'INFORMATION

Le besoin de disposer et d'analyser des données qui entretiennent des liens complexes dans la réalité aboutit actuellement à la constitution et à l'exploitation de nombreux **systèmes d'information.** Parmi ceux-ci, on peut citer les bases de données localisées qui se multiplient depuis plusieurs années; les enquêtes à objectifs multiples constituent une des sources d'information permettant d'enrichir ces systèmes d'information.

Il peut exister des liens de natures diverses entre les différentes informations composant les systèmes d'information, notamment des liens **hiérarchiques**. C'est le cas des enquêtes socio-économiques qui font intervenir différents niveaux de collecte et d'analyse tels que l'îlot, le logement, le ménage, la personne : une personne fait partie d'un ménage, qui est lui-même inclus dans un logement appartenant à un îlot donné. Lorsque ces différents niveaux sont emboîtés, on dispose d'un **système hiérarchique de l'information** ou encore un **système d'échelles.** Précisons que nous entendons par "échelle" le niveau d'analyse des données, qui "*ne se réfère pas nécessairement à un territoire circonscrit mais aussi éventuellement à un système de relations n'ayant pas toujours de traduction spatiale continue*"[1]. On peut attribuer à l'échelle une **dimension temporelle** ou **sociale,** comme par exemple pour l'étude des systèmes résidentiels. Si les individus ont une dimension **géographique,** on dispose alors un système d'informations localisées.

Un des principaux intérêts de tels systèmes d'information est de permettre de proposer une description de plus en plus riche de la réalité pour mettre à jour et exprimer formellement les interrelations existant entre les informations décrivant différents champs du réel.

Or, si les outils de traitement informatique des données (gestionnaires de base de données relationnelles, Systèmes d'Informations Géographiques...) ne cessent de se développer et permettent maintenant de gérer de manière organisée des ensembles complexes d'informations et de les exploiter, ils ne prennent pas toujours en compte le **fonctionnement** de ces systèmes sur lesquels repose pourtant l'analyse. La connaissance du fonctionnement d'un système d'information ne peut reposer que sur une étude approfondie des **liens entre les individus et entre leurs variables descriptives.** Pour cela, il est nécessaire de réfléchir sur **les outils de traitement statistique** qui permettent de décrire les systèmes d'informations, et de les maîtriser.

[1]J.C. BOYER dans [25ème Congrès, août 1984] p83.

OBJECTIF

Dans nombre de cas, les systèmes d'information répondent à des objectifs multiples afférant à des disciplines différentes. Le plus souvent, les données ne sont exploitées que de façon indépendante, en fonction de l'unité d'observation adaptée à l'approche spécifique de chaque discipline : peu de traitements conjoints mettant en relation plusieurs unités d'observation sont réalisés, sinon pour exprimer une relation entre deux variables. Il existe pourtant des besoins en matière de traitements, nécessitant d'avoir recours à des méthodes descriptives d'analyses multivariées. Si nous nous plaçons dans le cadre des bases de données localisées, trois types de traitements requièrent ainsi des méthodes d'analyses multivariées :

- le **traitement simultané d'un ensemble de variables qualitatives définies à des niveaux d'observation différents** : par exemple, les données relatives à l'îlot, au logement et au ménage recueillies, sur une même population ou un même échantillon, par une enquête socio-économique;
- **l'analyse conjointe de données issues de sources d'informations différentes** : on dispose souvent, pour un même espace, de données obtenues à partir d'enquêtes, de sources administratives ou autres, correspondant à des populations ou des échantillons différents. La mise en relation de ces informations issues de sources diverses suppose de déterminer le niveau d'analyse adéquat pour appréhender l'ensemble de ces données.
- **le croisement d'informations correspondant à des découpages spatiaux différents** : sur un même territoire, on dispose souvent de plusieurs sources d'information, pour lesquelles les unités spatiales de référence ne sont pas identiques, et pas nécessairement emboîtées. Diversité des thèmes et des sources d'information mobilisés dans les base de données localisées pose un problème récurrent dans l'analyse de ces données : le croisement d'informations dans un contexte de **recouvrement entre les unités spatiales** issues de ces découpages variables de l'espace étudié.

L'intérêt de l'exploitation des systèmes d'information repose sur **la mise en relation d'informations relatives à des niveaux différents.** L'analyse de tels systèmes peut permettre d'avancer dans la réflexion dans de nombreux domaines. Nous en citerons deux d'entre eux à titre d'exemple :

- **l'hétérogénéité spatiale** : très souvent, l'information traitée en géographie est observée à une échelle qui correspond à un certain niveau d'agrégation spatial. Il est même fréquent que cette information ait déjà été agrégée à plusieurs reprises et l'on dispose alors de plusieurs échelles emboitées. L'agrégation spatiale sur laquelle est fondée l'information géographique soulève des questions fondamentales pour le traitement et l'interprétation des analyses statistiques et cartographiques. En effet, **la nature, l'intensité et le sens des relations** entre les phénomènes

représentés par les variables **varient en fonction de l'échelle** à laquelle on se place : les agrégations *"affectent de manière fondamentale les caractéristiques des données, et donc les résultats de l'étude géographique"*[1].

Lorsque l'on n'a qu'un seul niveau d'analyse, le traitement des données agrégées pose clairement le problème de l'hétérogénéité spatiale. Si l'on dispose de plusieurs niveaux emboités, on peut envisager l'analyse multivariée de données agrégées.

- **la définition d'une stratification :** en amont du traitement de données d'enquête par sondage, se pose souvent le problème de la stratification de la base de sondage pour la sélection de l'échantillon d'enquête. Rappelons que la procédure de stratification consiste à *"constituer des groupes homogènes à l'intérieur de l'univers de l'étude avant de prélever l'échantillon afin d'amortir les fluctuations que le hasard peut apporter à sa composition"*[2]. Cependant, *"un mode de stratification peut se révéler très efficace pour la mesure d'une variable, très peu efficace pour la mesure d'une autre variable"*[3]. C'est pourquoi l'analyse d'un système d'information par les techniques d'analyses multivariées, **prenant en compte non pas une mais plusieurs variables,** peut se révéler d'un très grand intérêt pour l'élaboration d'une stratification.

L'objectif de notre travail est de proposer **des guides méthodologiques** pour fournir des éléments de réponse aux questions suivantes :

- Comment croiser des informations définies sur des niveaux différents ?
- Quelle est l'influence d'une échelle sur la perception d'un phénomène donné ?

Le principal problème réside dans la cohérence du traitement statistique mobilisant des informations relatives à des individus différents. En effet, si les outils informatiques permettent une manipulation aisée des données, il faut vérifier la pertinence des résultats de l'exploitation qui nécessite quelques règles bien précises : on ne croise pas des informations relatives à des individus de natures différentes sans une réflexion préalable, ni sans une transformation éventuelle (pondération, agrégation) des données.

C'est dans ce cadre de réflexion sur l'exploitation d'un ensemble d'informations résultant de plusieurs groupes de variables qualitatives décrivant des individus distincts, que nous proposons **une méthode d'analyse d'un système d'échelles à partir des techniques descriptives d'analyses multivariées.**

---

[1] [OPENSHAW, 1981]
[2] [DESABIE, 1966]
[3] [DESABIE, 1966]

DEMARCHE SUIVIE

Pour dégager les structures de l'univers composite que constitue le système d'information, **les méthodes d'analyse des données peuvent être des outils performants, permettant d'exploiter au mieux la structure organisée du système d'information.**

L'étude d'un tel système pose des problèmes de décomposition de ses éléments et de relations internes. Cela signifie d'un point de vue méthodologique qu'il faut explorer la structure hiérarchique de l'information pour construire un tableau cohérent qui respecte cette structure. Un certain nombre de choix doivent être faits puisque les informations sont définies sur des niveaux différents et donc caractérisent des populations différentes : on dispose de plusieurs populations statistiques et de plusieurs niveaux d'analyse potentiels.

Sur ces bases, on envisage alors **l'analyse statistique du système d'échelles** qui consiste à exploiter simultanément un ensemble de variables définies à plusieurs niveaux, sur autant d'échelles que l'étude le suggère. Il s'agit de mettre en évidence tant les structures caractéristiques de chaque niveau, du niveau individuel (micro) aux niveaux agrégés (macro), que les structures que l'on perd dans le passage d'un niveau à un autre. On observe ainsi les modifications de la perception d'un phénomène à chaque étape d'un changement d'échelle. Ceci nous conduit alors à envisager le suivi des différentes étapes par :

- **l'analyse par niveau d'agrégation** afin de cerner la structuration de l'univers d'étude sous des formes différentes ; chaque échelle fournit un point de vue particulier sur cette structuration. Ce type d'analyse a pour objet de mettre en évidence les différences entre les unités constituant chaque niveau d'agrégation étudié.

- **l'analyse du passage entre deux niveaux d'agrégation** afin d'établir la jonction entre ces différents points de vue. Cette analyse vise à observer **l'effet de l'agrégation,** c'est-à-dire la différenciation interne des unités constituant le niveau supérieur.

Par conséquent les questions méthodologiques que pose l'analyse statistique du système d'échelles sont à la fois **la mise en relation d'individus différents et l'analyse multivariée du changement d'échelles.**

PLAN DE L'EXPOSE

L'exposé est organisé en quatre parties. Les exemples d'application viennent en appui aux démonstrations et sont tirés du programme de recherche sur "*Les enjeux des extensions urbaines à Ouagadougou*"[1].

Dans la première partie, nous décrivons la construction d'une information hiérarchisée et nous fournissons quelques résultats obtenus par les tableaux croisés. Plusieurs unités d'observation ont été choisies : les parcelles, les ménages, les personnes composant les ménages, les itinéraires migratoires du ménage. Nous présentons également dans cette partie l'enquête socio-économique menée à Ouagadougou et les moyens mis en œuvre pour répondre aux objectifs du programme.

Dans la deuxième partie, nous proposons une réflexion sur la structure d'un tableau qui permette de traiter un ensemble de variables qualitatives définies sur différents niveaux d'observation. Il s'agit de concevoir un principe général pour la construction du tableau de données adapté à toutes les situations possibles qui résultent des choix à faire au moment du traitement statistique.

Dans la troisième partie, nous abordons l'analyse d'un système d'échelles. Il s'agit d'une part de déterminer les facteurs significatifs des différents niveaux et d'autre part d'étudier les effets d'un changement de niveau par la description du passage d'un niveau à un autre. Le principe est fondé sur la décomposition de l'inertie pour une série de partitions emboîtées.

Dans la quatrième partie, nous illustrons les développements méthodologiques précédents par une application aux données de l'enquête de Ouagadougou. On se propose d'étudier la structure de l'espace considéré comme produit des pratiques et des politiques urbaines de cette ville. Pour cela, une série d'analyses sont présentées pour mettre en évidence les formes d'organisation spatiale et sociale qui caractérisent la périphérie de cette agglomération africaine.

[1] [LE BRIS *et al*, 1992]

Première Partie

EXEMPLE D'ELABORATION D'UNE STRUCTURE HIERARCHIQUE DE L'INFORMATION

Les enquêtes qui font intervenir différents niveaux de collecte sont riches en information et permettent une meilleure approche d'une réalité complexe. Il s'agit alors dans le traitement des données de conserver le bénéfice de leur structure.

L'étude d'un système hiérarchique de l'information va être menée au travers d'une enquête sur les ménages résidant dans les quartiers périphériques de Ouagadougou constituant une des opérations d'un programme de recherche initié par l'Orstom et le Centre National de la Recherche Scientifique et Technique du Burkina Faso : "enjeux des extensions urbaines à Ouagadougou"[1].

Les objectifs de ce programme sont *"d'analyser le processus de production de la ville, des formes d'organisation spatiale et sociale qui la caractérisent à partir d'une analyse des pratiques et des stratégies des différents agents de cette production : la population citadine, les pouvoirs publics et locaux et les intervenants extérieurs. Il s'agit de mettre en rapport les desseins politiques sur la ville et les pratiques sociales et résidentielles des citadins"*. Il faut préciser que l'une des principales dimensions de la politique urbaine est le lotissement systématique des quartiers de la périphérie : certaines parcelles, voire certains quartiers ont été rasés.

C'est donc dans un contexte d'instabilité que s'inscrit l'enquête par sondage. *"Son objectif est de permettre une meilleure connaissance d'un phénomène mal appréhendé jusqu'ici en l'absence d'études à la fois extensives et approfondies : celui de la croissance des quartiers spontanés de la périphérie (son importance, ses modalités, ses causes, ses effets). Le souci d'une approche dynamique implique que l'on ne se contente pas uniquement d'enregistrer des données d'ordre synchronique (structures démographiques et socio-économiques, structures familiales et ethniques, origines géographiques, statuts fonciers et résidentiels), mais que l'on tente également d'obtenir des données d'ordre diachronique : les processus étudiés d'intégration à la ville sont étalés dans le temps, parfois sur plusieurs générations, et l'on doit prendre notamment en compte les itinéraires migratoires vers la ville et dans la ville, les modalités de l'installation et de l'accès au sol et au logement, les cursus professionnels, les remaniements de la structure familiale et matrimoniale, l'évolution des investissements fonciers et immobiliers, le phénomène de la mobilité intra-urbaine et de la dispersion résidentielle.(...)*

L'orientation de l'enquête explicitée nous a donc conduit, dans la mesure où nous avons pu la rendre compatible avec les nécessités propres à une exploitation informatisée, à introduire des questions d'ordre biographique portant non seulement sur le chef de ménage et l'accession à une parcelle, mais aussi, pour une part, sur l'ensemble du groupe familial. Ainsi avons-nous tenté de pallier à une lacune des enquêtes statistiques qui est d'abstraire l'individu de son environnement relationnel."[1]

[1] [LE BRIS *et al*, 1992]

Chapitre 1.1.

STRATEGIE D'ENQUETE

L'analyse des extensions urbaines de Ouagadougou impose une réflexion sur la stratégie d'enquête à adopter afin de répondre aux objectifs du programme de recherche : il s'agit d'étudier dans le temps non pas un mais plusieurs types d'individus statistiques, le ménage, les personnes le composant et ses itinéraires migratoires, sur un espace défini par la parcelle et pouvant être élargi à une unité spatiale plus grande. Ces types d'individus sont mis en relation entre eux par leur lieu de résidence ou leur lien de parenté. Cette stratégie d'enquête permet d'étudier un individu statistique en considérant le système social et spatial auquel il est appartient. Les implications en sont nombreuses et se situent à tous les stades de l'enquête : protocole des passages répétés, choix des unités d'observation et des unités d'analyse, définition de la (ou des) population statistique, plan de sondage, rédaction des questionnaires (rétrospectifs pour l'étude des biographies, répétés lors de plusieurs passages et répondant à différents cas de figures donnés par la mobilité des enquêtés), codification, saisie, domaines d'étude, méthodes d'analyse adaptées.

La méthodologie d'enquête doit tenir compte tant de la dimension temporelle de l'étude que de la structure de l'information qui implique l'existence de différents individus statistiques.

1. 1. 1. LA DIMENSION TEMPORELLE DE L'ENQUETE

Une des principales préoccupations de l'enquête est d'obtenir des données d'ordre diachronique. Deux techniques, déjà utilisées à de nombreuses reprises sur des terrains comme le nôtre[1], sont proposées pour approcher cette dynamique :

- l'enquête rétrospective qui permet d'observer les itinéraires migratoires du chef de ménage et par conséquent la mobilité intra-urbaine, l'intégration à la ville, le processus et les modalités de l'installation. Cette technique demande cependant, un effort de mémorisation de la part de l'enquêté[2].

[1][GODIN, 1987], [LE BRIS *et al*, 1977][INSEE-INED, 1977]
[2][COURGEAU, 1984]

- l'enquête à passages répétés qui permet non seulement d'étudier les mouvements démographiques antérieurs au 1er passage, mais aussi de cerner les changements récents intervenus entre les deux passages (mobilité des membres du ménage et mobilité des ménages). Cette méthode d'enquête nous permet ainsi d'appréhender le processus de croissance des quartiers spontanés, l'évolution des investissements fonciers et immobiliers, la dispersion résidentielle, phénomènes qui ne pourraient être que partiellement saisis par des enquêtes rétrospectives.

Pour la mise en place de l'enquête à passage répétés, on cherche dans la mesure du possible à maintenir un même échantillon aux différents passages. Cependant, si en milieu rural, ce type d'enquête, couramment utilisé, a fait ses preuves, en milieu urbain, il pose des problèmes aigus dans le choix de l'unité d'échantillonnage et d'enquête par le fait de la forte mobilité de la population. Une solution consiste à privilégier le suivi d'unités spatiales par rapport à celui des ménages. En effet, le ménage qui est un ensemble de personnes répondant à des critères sociologiques variables, est "*une unité statistique aux contours fluctuant dans le temps*"[1]. L'unité spatiale est donc théoriquement considérée comme plus stable.

1. 1. 2. DEFINITION DE LA POPULATION STATISTIQUE

Compte tenu des objectifs, on conçoit que plusieurs types de populations soient susceptibles d'être enquêtés. En effet, l'information doit être collectée au niveau :

- de la **parcelle** ou l'équivalent-parcelle : en milieu loti, la parcelle est une portion d'espace de forme généralement rectangulaire entourée d'une clôture. En milieu non loti, l'équivalent-parcelle, ou concession, peut être de forme irrégulière et délimité ou non par une clôture. Dans les deux cas, plusieurs ménages peuvent cohabiter.

- des **bâtiments** construits sur la parcelle.

- des **ménages** résidant sur la parcelle : tout couple marié ou tout individu isolé mais indépendant financièrement et sur le plan du logement.
  Afin de préserver la notion de groupe résidentiel, tous les ménages cohabitant sur une même parcelle ont été interviewés.

- de toutes les **personnes** composant le (ou les) ménage :

[1] J.C. DELVILLE dans [ASU, 1987] p92.

. enfant résidant ou non.

. épouse résidante ou non.

. dépendant : toute autre personne isolée non indépendante financièrement.

- des **migrations** du chef de ménage : une migration correspond à un changement de résidence.
- des **terrains** appartenant au chef de ménage : parcelle d'habitation ou champs de culture.

On précise qu'il n'y a pas de relation directe entre un ménage et un bâtiment.

Les types d'unités Parcelle, Bâtiment et Terrain permettent d'appréhender l'aspect foncier et immobilier. L'étude du Ménage fournit les caractéristiques socio-économiques de la population et complète les informations sur les modalités d'accès au sol et au logement. Les unités Epouse, Enfant et Dépendant permettent d'appréhender, d'une manière approfondie, la structure familiale et le fonctionnement interne du groupe familial. Enfin, l'enquête rétrospective permet de cerner les migrations (changement de résidence) aussi bien inter-urbaines qu'intra-urbaines.

Les enquêtes ont été mises en place de façon à effectuer des passages répétés sur des unités spatiales. Par conséquent, **la parcelle** a été retenue comme **unité d'échantillonnage**. Elle permet à elle seule d'atteindre hiérarchiquement les autres unités concernées et correspond à une unité physique, donc stable (*a priori*). Si l'information est centralisée au niveau de l'unité parcelle, elle est recueillie auprès des chefs de ménage c'est-à-dire de **l'unité ménage** qui constitue **l'unité d'enquête.**

La construction de types d'unités statistiques correspondant aux différents niveaux d'observation aboutit à une représentation arborescente de l'information schématisée par :

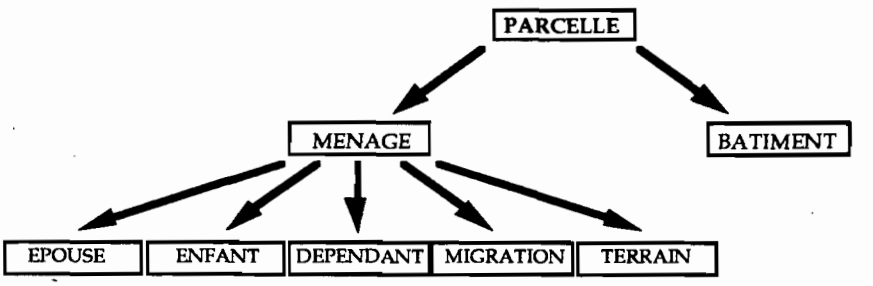


Figure n°1 : *Représentation hiérarchique de l'information*

L'existence de plusieurs types d'unités statistiques d'observation, Parcelle, Bâtiment, Ménage, Migration, Enfant, Epouse, Dépendant et Terrain **entre lesquelles il y a une relation d'inclusion définit une structure hiérarchique de l'information.**

PREPARATION DE LA COLLECTE ET DE L'ANALYSE DES DONNEES

Nous renvoyons aux annexes n°1.1 à 1.5 pour tout complément concernant la mise en place et la réalisation de l'enquête. Nous présentons ici uniquement les principes généraux de la collecte ainsi que les premiers résultats de l'enquête obtenus par les méthodes statistiques de base, tris à plat et tableaux croisés.

1. 2. 1. PLAN DE SONDAGE

Les principales difficultés rencontrées pour la mise en place de l'enquête étant dues au terrain d'étude, nous nous sommes appuyés sur des ouvrages concernant les enquêtes dans les pays en voie de développement[1].

L'impératif de représentativité de l'échantillon est d'exprimer toute la population périphérique de Ouagadougou, aussi hétérogène soit-elle. Aucune zone spatiale ne doit être privilégiée. Il s'agit donc d'identifier chacune d'elles, puis au moment de la constitution de l'échantillon, de respecter au mieux la répartition géographique de la population.

Par ailleurs, la conception même de l'enquête introduit la notion d'emboîtement des populations statistiques ; un secteur définit un ensemble de parcelles ou de ménages ou d'individus ; une parcelle définit un ensemble de ménages ou d'individus ; un ménage définit un ensemble d'individus. Elle implique l'existence de différents degrés dans le plan de sondage. Comme l'écrit J.C Delville[2], un plan de sondage à plusieurs degrés est "*une suite emboîtée de tirages à deux degrés. Au plan théorique, les choses se ramènent à la définition d'une partition sur l'univers à sonder....On utilise* (pour un tirage à degrés multiples) *une suite emboîtée de partitions de plus en plus fines définissant successivement les unités primaires puis secondaires. Il s'agit en d'autres termes, d'une classification hiérarchique*". Ainsi, les populations des niveaux supérieurs engendrent une ou plusieurs partitions sur les populations des niveaux inférieurs et ordonnés dans la hiérarchie.

[1] [ASSELIN, 1984]; [DUREAU, 1989]; [GODIN, 1987]
[2] [ASU, 1987] pp106-107.

Le tableau ci-dessous présente les caractéristiques, qui sont détaillées en annexe 1.1, du plan de sondage adopté :

Structure générale	**Sondage stratifié en grappe :** 1er degré = parcelles habitées (grappes) 2ème degré = ménages (unités statistiques)
Base de sondage	Ensemble des parcelles habitées de la périphérie
Stratification	Strates = secteurs du recensement de la population de 1985
Sélection des parcelles (1er degré)	Tirages systématiques indépendants au 1/43 sur chaque strate **Unités d'échantillonnage**
Sélection des ménages (2ème degré)	Recensement dans chaque parcelle **Unités d'enquête**
Estimateur	Proportions

Tableau n°1 : *Caractéristiques du plan de sondage*

1. 2. 2. LE QUESTIONNAIRE

La structure de l'information fait intervenir différents types d'unités d'enquête. Il s'agit constamment de replacer chaque unité statistique (parcelle, bâtiment, ménage, individu composant le ménage ou migration du chef de ménage) dans un contexte plus large (en relation avec les autres unités). Pour cela, sachant qu'un questionnaire est un ensemble de questions se répétant identiquement pour chaque unité statistique, cette enquête peut être envisagée sous l'aspect d'un ensemble de sous-questionnaires, eux-mêmes relatifs à un ensemble de sous-unités statistiques, composant ainsi le questionnaire de base.

Le questionnaire touche différents thèmes liés aux objectifs: thèmes relatifs au foncier, à la mobilité et à la structure familiale, complétés par les questions d'identification de la population (caractéristiques sociologiques, démographiques et économiques), et étudiés dans leur évolution à un an d'intervalle.

- L'étude du foncier est axée essentiellement sur l'accession au sol (mode d'acquisition de la parcelle, dates, opérations de lotissement) en référence à la parcelle, mais aussi au statut résidentiel du chef de ménage et son évolution.

- L'analyse de la mobilité est envisagée à la fois par l'enquête rétrospective afin de cibler les itinéraires vers et dans la ville et par le deuxième passage afin de mieux cerner la mobilité des ménages et de la population résidant en périphérie.

- La structure familiale est sans doute la plus difficile à cerner car elle concerne différents types

de familles et fait intervenir les relations de parenté entre d'une part les individus composant le ménage et d'autre part les ménages co-résidents.

Plusieurs types de questionnaires ont été conçus en fonction du passage de l'enquête et des différentes situations rencontrées et sont présentés en annexes n°1.2 et n°4. Rappelons que les informations concernant aussi bien la parcelle, les bâtiments, les individus du ménage que le ménage ont été recueillies auprès du chef de ménage .

1. 2. 3. LES ANALYSES BIVARIEES

Les analyses statistiques sont abordées par l'intermédiaire de tableaux d'informations. Chacun d'entre eux est construit par rapport à un type d'unité statistique : on doit travailler sur une population statistique homogène, les parcelles, les ménages ou les individus.... Or nous avons autant de populations statistiques qu'il y a de types d'unités observés. Par conséquent, pour chaque traitement, il doit être précisé obligatoirement, la population statistique d'étude dépendant du type d'unité [1] (conséquence de la structure hiérarchique) mais aussi le passage d'enquête [2] (étude diachronique) et d'une manière facultative, la (ou les) caractéristique(s) spécifique(s) du domaine d'étude [3] (statut d'occupation, tissu urbain, groupe résidentiel, variables permettant d'approfondir un thème) en ayant soin de vérifier que l'effectif de la sous-population de travail résultant de tous ces critères soit suffisamment élevé (supérieur à 30) pour que les calculs de proportions aient un sens.

Les exemples suivants illustrent la démarche pour approcher la structure hiérarchique et les deux passages d'enquête :

[1]		Etude sur la population Ménage		
[2]		1er passage	2ème passage	Ensemble des ménages enquêtés
[3]	Propriétaire	603 (+10)	518	627
	Locataire	76	73	99
	Hébergé	246 (+3)	223	294

[1]		Effectif parcelle	Effectif ménage
[2] et [3]	Zone lotie avant le 1er passage	180	224
	Zone lotie entre les 2 passages	292	336
	Zone en cours de lotissement	229	274
	Zone non lotie	139	186

[1]		Effectif parcelle		Effectif ménage	
[2]		1er passage	2ème passage	1er passage	2ème passage
[3]	Groupe monorésidentiel	777 (+13)	623	777 (+13)	623
	Groupe polyrésidentiel	63	81	148	191
	Parc. inoccupée (2ème passage)	-	136	-	-

Les douze tris croisés et tris à plats présentés et commentés ci-dessous proviennent du rapport Orstom sur les "Enjeux des extensions urbaines à Ouagadougou". La moyenne et le total sont donnés pour chaque tableau et indique leur lecture en lignes et/ou en colonnes (*italique*). Ils ont été sélectionnés, d'une part pour montrer comment les méthodes simples de traitement de données ont été utilisées pour appréhender la structure hiérarchique de l'information et l'enquête à passages répétés et, d'autre part pour donner un bref aperçu des premiers résultats de l'enquête. Ceux-ci montrent l'originalité de Ouagadougou et permettent de mettre en évidence les thèmes dominants de l'étude qui, dans cette présentation, reposent sur 6 variables "clés" :

- structure familiale

Etude de la structure du groupe résidentiel

FAMILLES ELEMENTAIRES				MENAGES DE "CELIBATAIRES"				GROUPE POLY-NUCLEAIRES	Total
Restreintes		Elargies		Hommes		Femmes			
31		39		16		7		7	100
mono-gamique	poly-gamique	mono-gamique	poly-gamique	non mariés	mariés	non mariées	mariées		
23	8	28	11	14	2	4	3	7	100

Population : PARCELLE / 1er passage

La forte proportion des parcelles occupées par un seul ménage, situation peu commune en ville africaine, peut s'expliquer de trois manières :
. la jeunesse du phénomène du peuplement des quartiers périphériques ;
. la faible superficie des parcelles;
. la politique de lotissement systématique, "à chaque ménage, sa parcelle", qui favorise un éclatement des ménages dans le but de se faire attribuer le plus de parcelles possible. Ainsi s'expliquent également le grand nombre de ménages "célibataires" et la prédominance des structures du types famille élémentaire par "la stratégie délibérée de dispersion".

- foncier:

Etude de la superficie des parcelles selon son mode d'obtention

Superficie	Mode d'obtention de la parcelle					Moyenne
	Attribution par l'Etat	Achat	Attribution coutumière	Héritage	Don	
< 250 m2	4	17	17	12	28	18
250 - 349 m2	50	22	26	28	19	27
350 - 449 m2	32	21	9	22	23	23
450 - 749 m2	12	27	20	22	18	20
> 750 m2	2	12	28	15	13	12
Total	100	100	100	100	100	100
Moyenne	26	28	16	8	22	

Population : PARCELLE / 1er passage

Il ressort un contraste très net entre les modes d'obtention "traditionnels" et "l'attribution par l'Etat" qui influe sur la superficie de la parcelle.

Grille de peuplement

Nombre de personnes par parcelle	Nombre de pièces par parcelle						
	1	2	3	4	5	6	7 et +
1							
2 - 4				sous-peuplement			22%
5 - 7			peuplement normal 66%				
8 et +	sur-peuplement			12%			

Moyenne: 6,1 personnes par parcelle
1,7 personnes par pièce
3,6 pièces par parcelle

Population : BATIMENT / 1er passage

12% des personnes vivent en situation de sous-peuplement. Si l'on tient compte du fait que les constructions, majoritairement en banco dans la périphérie, sont réalisables en peu de temps, on constate malgré tout un décalage entre la croissance de la famille et celle de l'espace habité.

- La mobilité entre les deux passages :

La mobilité a affecté en un an plus du quart des enquêtés. Il s'agit donc d'un phénomène important. Nous allons le présenter au travers :

- du foncier :

Mobilité des ménages entre les deux passages et opérations de lotissement

	Ménages partis	Ménages stables	Ménages arrivés	Total	Moyenne
Lotie avant le 1er passage	8 *9*	84 *25*	8 *21*	100	*22*
Lotie entre les deux passages	36 *61*	58 *26*	6 *23*	100	*32*
En cours de lotissement au 2ème passage	9 *13*	82 *31*	9 *29*	100	*27*
Non lotie	17 *17*	71 *18*	12 *27*	100	*18*
Total Moyenne	*100* 19	*100* 73	*100* 8	 100	*100*

Population: MENAGE / 2ème passage
La variable "opération de lotissement" est une variable dérivée de deux variables de base, chacune spécifique d'un passage.

La politique urbaine influe sur la mobilité des ménages en un an. On constate au 2ème passage que le lotissement a provoqué un départ des ménages dû au phénomène de déguerpissement.

Statut d'occupation et mobilité entre les deux passages

	Propriétaire	Locataire	Hébergé	Total	Moyenne
Ménages partis	54 *16*	11 *22*	35 *23*	100	*19*
Ménages stables	69 *82*	7 *60*	24 *61*	100	*73*
Ménages arrivés	17 *2*	24 *18*	59 *16*	100	*8*
Total Moyenne	*100* 62	*100* 10	*100* 28	 100	*100*
1er passage	66	8	26	100	
2ème passage	64	9	27	100	

Population:MENAGE /2ème passage

A peine 1% des ménages non migrants ont changé de statut d'occupation entre les deux passages. De ce fait, le statut d'occupation des ménages stables correspond à celui observé lors du 1er passage de l'enquête.
La mobilité entre les deux passages affecte les ménages locataires et hébergés. On relève également l'absence de propriétaire parmi les ménages arrivés.

Motifs de départs ou d'arrivée

	Motifs				Total
	fonciers	économiques	sociaux	non-précisé	
Ménages partis	60	18	15	7	100
Ménages arrivés	56	14	26	4	100

Population : MENAGES MIGRANTS / 2ème passage

Les motifs fonciers sont les raisons principalement évoquées pour la mobilité des ménages.

Mobilité interne des ménages non migrants

	Ménages avec mobilité interne			Ménages sans mobilité interne	Total
	Solde migratoire : SM				
	SM<0	SM=0	SM>0		
Lotie avant le 1er passage	20	6	28	46	100
Lotie entre les 2 passages	23	3	18	56	100
En cours de lotissement	16	1	21	62	100
Non lotie	22	6	24	48	100
Ensemble périphérie	20	3	23	54	100

Population : MENAGES NON MIGRANTS / 2 passages
Les variables "opération de lotissement" et "solde migratoire" sont des variables dérivées et résultantes des deux passages d'enquête.

La mobilité interne est un phénomène mineur. Plus de la moitié des ménages considérés ne connaissent aucun mouvement au moment précis des deux passages (mais les entrées et les sorties au cours de la période de référence n'ont pas été enregistrées). L'opération de lotissement (loti entre les deux passages et en cours de lotissement) semble avoir eu un effet de ralentissement de la mobilité au sein des ménages. Cependant, elle provoque des départs d'individus plus accentués. On conclut malgré tout qu'il n'y a pas de relation évidente entre la mobilité interne et les opérations de lotissement.

- structure économique :

Secteurs d'activité mobilité entre les deux passages

	Secteur agricole	Secteur informel	Secteur moderne	Non actif	Total	Moyenne
Ménages partis	14	33	40	13	100	
	17	*19*	*18*	*28*		*19*
Ménages stables	17	33	43	7	100	
	76	*73*	*74*	*60*		*73*
Ménages arrivés	15	29	43	13	100	
	7	*8*	*8*	*12*		*8*
Total	*100*	*100*	*100*	*100*		*100*
Moyenne	16	32	43	9	100	
1er passage	16	33	42	9	100	
2ème passage	16	32	43	9	100	

Population : MENAGE / 1er et 2ème passages

L'activité mentionnée est celle que le chef de ménage non migrant exerçait au 1er passage. En effet, 1,5% des chefs de ménages non migrants ont changé de secteurs d'activité soit 2% n'exercent plus la même activité entre les deux passages parmi ces ménages. Entre les deux passages, les variations sont infimes. Mais rappelons que les deux passages d'enquête ont été réalisés à un an d'intervalle pour éliminer les variations saisonnières. On note que les fluctuations se sont faites parmi les chefs de ménages non-actifs.

Activité de l'épouse résidente selon l'activité du chef de ménage

Secteur d'activité du chef de ménage	Secteur d'activité de l'épouse sect. informel	sect. moderne	Epouses non actives	Total	Moyenne
Secteur informel	31	1	68	100	55
Secteur moderne	16	9	75	100	42
Non productif	20	0	80	100	3
Moyenne	25	4	71	100	

Population : EPOUSE RESIDENTE / 1er passage

La plupart des épouses résidant avec leur mari n'ont pas d'activité professionnelle "déclarée". Proportionnellement, elles exercent une activité dans le même secteur que leur mari.

- intégration à la ville :

Itinéraires migratoires

Nombre de migrations	Types d'itinéraires migratoires pas de migration Hors de Ouagadougou		une ou plusieurs migrations Hors de Ouagadougou		Total	Moyenne
	Ouagadougou	Hors Ouaga .- Ouagadougou	Ouaga. - H.O - Ouaga.	H.O - Ouaga. - H.O - Ouaga.		
Aucune migr.	100	-	-	-	100	3
1 migration	26	74	-	-	100	23
2 migrations	6	88	6	-	100	34
3 migrations	1	89	5	5	100	19
4 migrations	10	70	3	27	100	10
5 migrations	0	52	13	35	100	4
6 et + de migr.	0	41	8	51	100	7
Moyenne	11	76	4	9	100	

Population : MENAGE / 1er passage

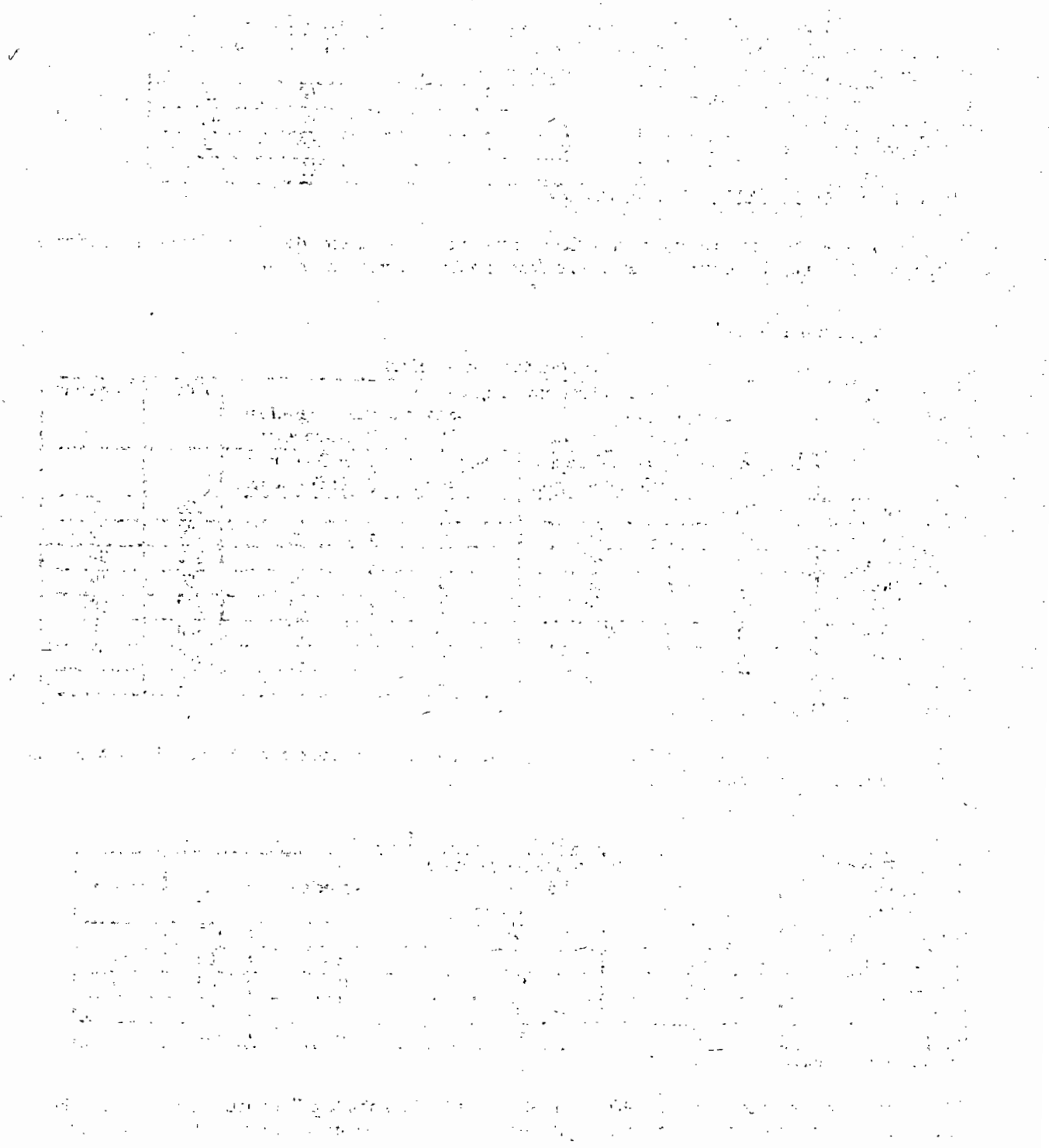
L'immigration directe du lieu de naissance à la capitale et ayant connu une étape intermédiaire l'emporte sur les itinéraires plus complexes.

Migrations intra-urbaines

Nombre de migrations intra-urbaines	Itinéraires intra-urbains Centre - Périphérie	Périphérie - Périphérie	non-précisé	Total	Moyenne
aucune migration	-	100	-	100	40
1 migration	65	28	6	100	46
2 migrations	80	10	10	100	10
3 ou +	75	10	15	100	4
Moyenne	41	54	5	100	

Population : MENAGE / 1er passage

Là encore, les chefs de ménage enquêtés manifestent une faible mobilité à l'intérieur de la ville. L'arrivée à Ouagadougou s'est faite à peu près à parts égales entre les quartiers centraux et les quartiers périphériques.



VERS L'EXPLOITATION DE LA STRUCTURE HIERARCHIQUE DE L'INFORMATION

La structure hiérarchique de l'information permet de lire progressivement les informations depuis celles relatives à la parcelle vers celles caractérisant les individus ou les migrations, et réciproquement. Cette conception de l'information offre l'avantage de traiter une catégorie d'enquête qui sera plus riche que les enquêtes restreintes à un seul type d'unité qui agrègent et donc perdent l'information des sous-unités.

Une variable définie au niveau Ménage peut être perçue à son propre niveau, c'est-à-dire celui du Ménage, mais aussi au niveau supérieur, celui de la Parcelle, ou au niveau inférieur, celui de l'Individu. Le croisement de deux variables (ou plus) prises à des niveaux différents et caractérisant deux populations (ou plus) statistiques différentes n'est pas immédiat et impose des choix à faire sur l'une de ces populations. Si on considère une variable Parcelle et une variable Ménage, le croisement peut être appréhendé de deux manières :

- sur la population Parcelle (correspondant à l'unité supérieure imposée par les variables concernées), ce qui implique que l'on sélectionne une modalité de la variable qualitative du niveau le plus fin et qui devient une variable à part entière au niveau supérieur ;

- sur la population Ménage (correspondant à l'unité inférieure imposée par les variables à traiter), où la valeur de la variable Parcelle est répétée autant de fois qu'il y a de ménages sur la parcelle. On s'intéresse alors aux interactions entre les variables en créant délibérément une répétition de l'information des unités supérieures. De ce fait, cette information est appréhendée sur une population statistique différente de celle d'origine.

Le dépouillement des données d'enquêtes qui abordent plusieurs niveaux de collecte peut ainsi se concevoir soit au niveau le plus fin sur la population statistique correspondante, soit à un niveau supérieur impliquant une réduction de l'information contenue dans les niveaux plus fins par la création de nouvelles variables. N'exploitant pas la structure hiérarchique de l'information, ces procédures entraînent bien évidemment une réduction importante des données : on perd en effet l'environnement relationnel des individus, autorisé par une collecte à plusieurs niveaux.

La connaissance du fonctionnement du système dans sa globalité passe par l'étude du lien entre les niveaux, jusqu'à présent rarement exploité explicitement. Nous proposons d'exploiter la structure

hiérarchique de l'information en utilisant l'analyse des correspondances appropriée au traitement de variables qualitatives. Nous allons nous appuyer pour cette étude sur des exemples issus de l'enquête où nous considérerons l'emboîtement Personne-Ménage-Parcelle-Secteur avec les niveaux Personne, Ménage et Parcelle qui sont des niveaux d'observation.

Deuxième Partie

CONCEPTION DU TABLEAU POUR EXPLOITER UNE STRUCTURE HIERARCHIQUE DE L'INFORMATION

La mise au point d'un système de pondération sur toutes les unités statistiques est un moyen pour construire un tableau qui croise des variables qualitatives définies aux différents niveaux d'observation et qui respecte la structure hiérarchique de l'information.

Le lien hiérarchique peut être vu comme un outil conceptuel très opératoire pour décrire une réalité complexe à travers un ensemble d'informations structuré et organisé, en permettant une approche à la fois synthétique, multidimensionnelle et pourquoi pas systémique.

La structure arborescente conduit d'une part à analyser une même information à différents niveaux, d'autre part à procéder à des analyses combinant des informations correspondant à différents niveaux d'observation.

Les questions soulevées par l'exploitation d'un tel ensemble de données sont de deux ordres :

- Comment croiser des variables qui se trouvent à des niveaux différents, emboîtés ou non, et qui correspondent alors à des populations statistiques différentes ?
- Comment appréhender un ensemble de variables sur différents niveaux ?

L'analyse multivariée constitue un outil de grand intérêt pour traiter ces questions. La réflexion réside alors dans la construction d'un tableau **rectangulaire**, constitué de **données homogènes** et respectant la **structure hiérarchique**.

Chapitre 2.1

DES DONNEES AUX TABLEAUX

On cherche à concevoir un tableau d'information qui permet de croiser les variables qualitatives définies à des niveaux d'observation différents. Le problème est donc de transcrire l'information ainsi structurée sous la forme d'un tableau rectangulaire de données homogènes.

2. 1. 1. QUELQUES DEFINITIONS ET RAPPELS

Il est nécessaire auparavant d'introduire quelques définitions pour formaliser le concept de la structure hiérarchique de l'information ce qui est indispensable à la mise en place du tableau d'information.

- "*Une **unité statistique** est un élément d'un ensemble* (ensemble I) *sur lequel sont effectuées des statistiques*"[1] ; on distingue :
  - les **unités d'observation** qui correspondent également aux individus statistiques dont l'ensemble définit la population statistique,
  - "*les **unités d'analyse** qui ne sont pas forcément observées, bien que pouvant l'être*". Ses caractéristiques sont déduites de celles des unités d'observation.

- Un **type d'unité** est une catégorie d'unités statistiques : le Secteur, la Parcelle, le Ménage, le Bâtiment, la Personne ....
  Un type d'unité d'observation correspond à un **niveau d'observation** (la Parcelle, le Ménage, le Bâtiment, la Personne) et un type d'unité d'analyse définit un **niveau d'analyse** (le Secteur, la Parcelle, le Ménage, le Bâtiment, la Personne).
  On distingue :
  - les niveaux ordonnés dans la hiérarchie c'est-à-dire les **niveaux emboîtés** : par exemple Parcelle-Ménage-Personne,
  - les **niveaux indépendants** : par exemple Ménage-Bâtiment ou Personne-Bâtiment.

[1] [DUBOIS *et al*, 1989]

On rappelle alors que l'existence de plusieurs types d'unités statistiques d'observation emboîtés définit une **structure hiérarchique de l'information.**

- Une **population statistique** est l'ensemble des individus statistiques c'est-à-dire des unités d'observation pour un type d'unité donné.

- "*Les **variables** ,* ou encore attributs, (ensemble J) *expriment les caractéristiques des unités statistiques relevées sous une forme qualitative ou quantitative.*"[1]
  Chaque type d'unité d'observation est décrit par un groupe de variables.

- "*La **donnée** est l'unité élémentaire d'information*" et correspond à la valeur d'une variable correspondant à une unité statistique.

"*Un **tableau d'information** est la forme à deux dimensions la plus simple qui permette de représenter une masse d'information à l'état brut. Il comprend : en ligne les unités statistiques, en colonnes les variables, et dans les cases, à l'intersection des lignes et des colonnes, les données*"[1] homogènes c'est-à-dire qui concernent les mêmes individus statistiques. On note K_{IJ} le tableau où I est l'ensemble des unités statistiques i et J est l'ensemble des variables j. K_{IJ} est défini sur le produit des deux ensembles finis I*J qui donne la dimension du tableau d'information.

Il est alors habituel de décrire un ensemble de variables qualitatives (des données d'enquête par exemple) par l'analyse des correspondances multiples, celle-ci étant une généralisation de l'analyse de tableaux croisés. On obtient les tableaux qui croisent les variables entre elles à partir de tableaux qui juxtaposent les variables. On aboutit ainsi à deux types de tableaux d'information traitant plusieurs variables qualitatives :

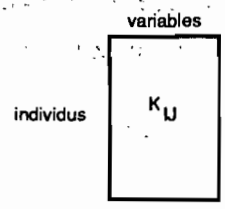


Tableau juxtaposant les variables

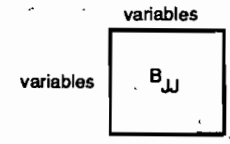


Tableau croisant les variables

- **le tableau (individus, variables)** défini sur I*J (ou encore (unités, variables) puisque les individus correspondent à des unités d'observation), qui juxtapose les variables tel que le tableau disjonctif complet où $K_{ij} = 1$ ou 0 selon que l'individu statistique i a adopté ou non la modalité j.

- **le tableau (variables, variables)** défini sur J*J qui croise toutes les variables entre elles, appelé également **tableau de Burt.** Il s'agit d'une juxtaposition de tous les tableaux croisés.

[1][DUBOIS *et al*, 1989]

On passe du premier au deuxième par la formule : $B_{jj'} = \sum_i \frac{K_{ij} K_{ij'}}{K_i}$,

et l'on comptabilise ainsi, lorsque K_{IJ} est un tableau disjonctif complet, le nombre d'individus qui ont répondu à la fois à la variable en ligne et à la variable en colonne et pondéré par le coefficient 1/k (où k est le nombre de variables).

Par exemple si 6 personnes ont été interrogées sur leur secteur d'activité et leur statut résidentiel, l'effectif résultant du croisement de ces deux variables s'obtient au coefficient 1/2 près de la manière suivante :

Individus \ Variables	S.informel	S.moderne	Sans prof.	prop.	loc.	heb.
1	1	0	0	0	1	0
2	0	1	0	1	0	0
3	0	1	0	0	0	1
4	0	0	1	0	0	1
5	1	0	0	1	0	0
6	0	1	0	1	0	0

Variables \ Variables	S.inf.	S.mod.	S prof.	prop.	loc.	heb.
S.inf.	2	0	0	1	1	0
S.mod.	0	3	0	2	0	1
S prof.	0	0	1	0	0	1
prop.	1	2	0	3	0	0
loc.	1	0	0	0	1	0
heb	0	1	1	0	0	2

Figure n°2 : *Construction d'un tableau de Burt*

Remarques :

- le tableau de contingence, couramment appelé tableau croisé, est le tableau classique qui permet d'exprimer la liaison entre deux variables qualitatives. Il s'agit d'un tableau d'effectifs obtenus en croisant les modalités des deux variables définies sur une même population d'individus. Il correspond à la fois :
  - au type de tableau (variable, variable) puisque c'est le tableau élémentaire qui compose le tableau de Burt;
  - au type de tableau (unités, variables) où il s'agit non pas d'unités d'observation mais des unités d'analyse.

- Le tableau d'information traitant de variables qualitatives, le tableau de contingence et par extension le tableau de Burt, est pertinent si **les marges,** c'est-à-dire les sommes en ligne et en colonne, **ont un sens** ; le plus souvent, elles expriment des effectifs définis sur une même population à un coefficient près.

- On rappelle également que les analyses des correspondances réalisées sur un tableau qui juxtapose les variables et celle effectuée sur le tableau associé qui croise les variables sont équivalentes : on obtient les mêmes facteurs bien que les valeurs d'inertie soient différentes ; les individus dans l'analyse du tableau de Burt sont sous-jacents.

2. 1. 2. LES DIVERSES DEMARCHES POUR STRUCTURER LE TABLEAU D'INFORMATION

La structure hiérarchique de l'information, représentée par la figure n°3, n'est donc pas conforme aux structures rectangulaires usuelles imposées par les méthodes d'analyses multidimensionnelles telle que l'analyse des correspondances à laquelle nous nous attachons ici et qui est appropriée au traitement de variables qualitatives.

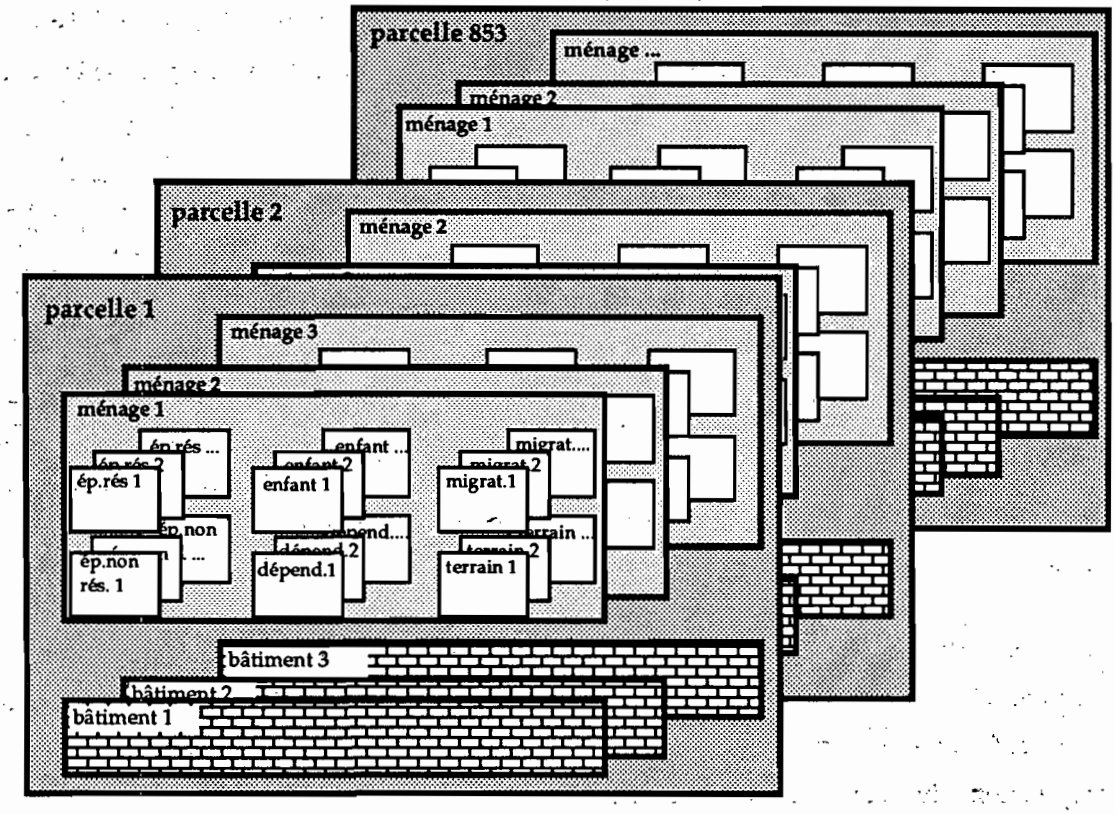


Figure n°3 : *Structure hiérarchique de l'information*

Deux démarches sont dans un premier temps considérées pour ne constituer qu'un seul tableau, par exemple:

- la première démarche consiste à définir un tableau sur la base des unités statistiques correspondant au niveau supérieur, la Parcelle, et toutes les autres unités les caractérisant (bâtiments, ménages, individus, migrations) sont considérées ; ce qui oblige à prévoir le nombre maximum d'unités rencontrées pour chaque type d'unités.

parc. 1 33 var.	bât. 1 14 var.	bât. 2 14 var.	...	bât. b 14 var.	men.1 150 var.	...	men. m 150 var.	enf. 1 8 var.	...	enf. e 8 var.	..
parc. 2 33 var.	bât. 1 14 var.	bât. 2 14 var.	...	bât. b 14 var.	men.1 150 var.	...	men. m 150 var.	enf. 1 8 var.	...	enf. e 8 var.	..

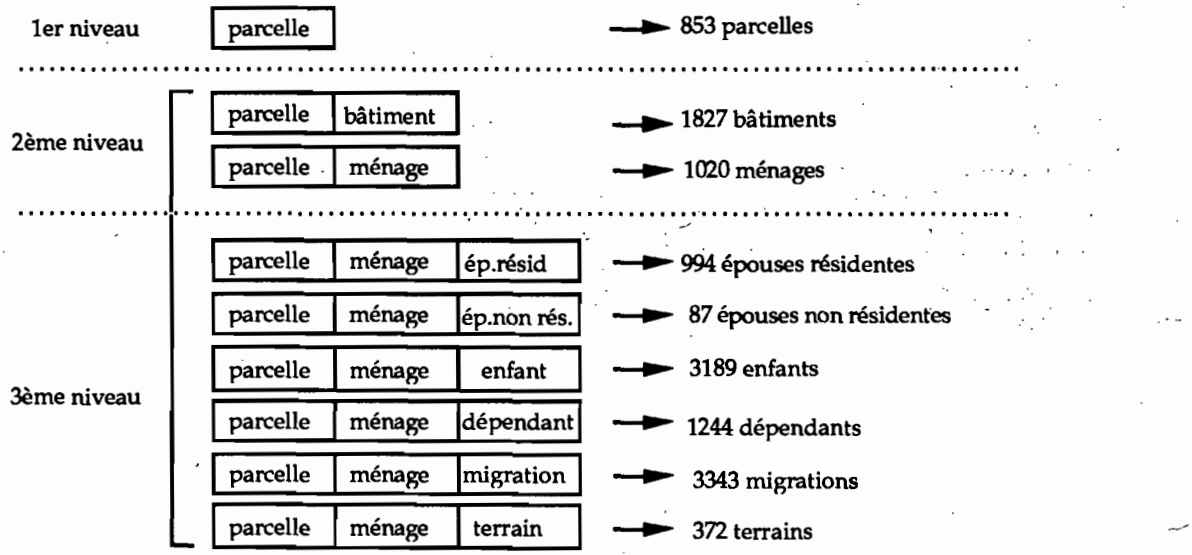
où b est le nombre maximum de bâtiments, m le nombre maximum de ménages, e le nombre maximum d'enfants, ..., rencontrés sur une parcelle.

- une deuxième démarche consiste à concevoir un tableau basé sur les unités statistiques inférieures dans lequel les unités supérieures et parallèles sont répétées autant de fois qu'il est nécessaire.

parcelle 1	bâtiment 1	ménage 1	enfant 1	dépendant 1	...
				dépendant 2	
				dépendant 3	
			enfant 2	dépendant 1	
				dépendant 2	
				dépendant 3	
		ménage 2	enfant 1	dépendant 1	
				dépendant 2	
	bâtiment 2	ménage 1	enfant 1	dépendant 1	
...	...				
parcelle 2	bâtiment 1	ménage 1	enfant 1	dépendant 1	...
...	...				

Ces deux démarches fastidieuses ne sont ni satisfaisantes ni praticables : le nombre de rubriques est trop important et la répétition des observations ou des unités n'est pas justifiée.

Une troisième démarche consiste à construire non pas un tableau mais plusieurs sous-tableaux, en l'occurrence 9, qui respectent la hiérarchie:



Cette décomposition ne permet pas d'étudier directement des types d'unités contenus à des niveaux indépendants dans la hiérarchie (les bâtiments avec les ménages). Elle offre cependant la possibilité de traiter ensemble des unités qui se succèdent immédiatement dans la hiérarchie (parcelles, ménages, dépendants). Pour cela, l'information parcelle (respectivement ménage) est répétée autant de fois qu'il y a de ménages sur la parcelle (respectivement de personnes dans le ménages).

L'approche immédiate du tableau d'information est, en fait, une décomposition en plusieurs sous-tableaux rectangulaires qui intègrent progressivement un niveau, deux niveaux et trois niveaux

consécutifs d'observation. Ils diffèrent par la nature de leurs populations, du nombre d'individus statistiques et des groupes de variables : pour chaque sous-tableau, le niveau le plus fin détermine la population d'étude et l'espace des individus ; les groupes de variables s'accumulent en fonction de l'intégration des niveaux. Il y a autant de tableaux d'information différents qu'il y a de types d'unités statistiques distincts.

Nous allons maintenant formaliser la démarche pour construire un ou plusieurs tableaux qui abordent une même population statistique suivant différentes échelles de perception et qui juxtaposent ou croisent l'ensemble des variables définies à plusieurs niveaux.

PRINCIPE DE CONSTRUCTION DU TABLEAU D'INFORMATION

La cohérence des résultats dépend de la pertinence du tableau d'information qui est soumis à l'analyse statistique.

2. 2. 1. LES CONDITIONS

On rappelle que, pour le traitement de variables qualitatives, les contraintes imposées par le tableau de l'information sont :

- une population statistique homogène : on travaille sur les mêmes individus statistiques,
- un tableau rectangulaire : les lignes représentent les unités statistiques d'analyse (qui ne sont pas forcément les individus statistiques) et les colonnes les variables.

Ces deux conditions réunies permettent au tableau d'être pertinent c'est-à-dire d'avoir des marges cohérentes. On conçoit alors que la structure hiérarchique ne permet pas d'emblée la construction d'un tel tableau puisqu'elle est définie par l'existence de plusieurs types d'unités d'observation (c'est-à-dire de plusieurs niveaux d'observation) emboîtés ce qui implique l'existence de :

- **plusieurs populations statistiques potentielles** : il y a autant de populations statistiques qu'il y a de types d'unités d'observation. On compte des personnes ou des ménages ou des parcelles.
- **plusieurs niveaux d'analyse** qui sont déduits des niveaux d'observation : chaque niveau d'observation est lui-même un niveau d'analyse, mais un niveau d'analyse peut ne pas être directement observable.
- **plusieurs groupes de variables** descriptives des populations et donc des niveaux d'observation : on dispose de variables descriptives de la Personne, du Ménage et de la Parcelle.

Par conséquent, il convient de se demander constamment :

- sur quelle population,
- et sur quel niveau d'analyse,

l'analyse bivariée ou multivariée doit être effectuée. Par exemple, si on choisit de travailler sur la population des ménages (on peut également travailler sur l'ensemble des personnes ou des parcelles), on cherche à exploiter les variables définies à tous les niveaux (Parcelle, Ménage et Personne) soit à l'échelle du Ménage ou à celle de la Parcelle ou de la Personne mais aussi à une échelle plus globale, le Secteur par exemple.

Quelques soient les choix à faire, il s'agit d'**élaborer un principe général pour la construction du tableau d'information** qui juxtapose (ou qui croise) l'ensemble des variables définies à plusieurs niveaux.

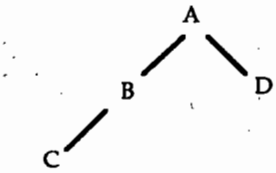
Nous avons vu alors que la structure hiérarchique de l'information ne permet pas d'étudier directement la relation qui peut exister entre deux informations qui appartiennent à des unités correspondant à des niveaux indépendants. Par contre, les informations contenues dans un type d'unité donné peuvent être mises immédiatement en relation soit entre elles, soit avec les informations des unités appartenant aux autres niveaux emboîtés dans la hiérarchie.

Nous allons considérer les deux situations suivantes :

niveaux emboîtés — niveaux indépendants

Quels sont les tableaux que l'on peut construire sachant que généraliser ou restreindre à partir de ces deux situations ne pose pas de problème ?

On peut alors aborder une structure plus complexe du type :



les différents types de niveaux, emboîtés et indépendants

2. 2. 2. LA RELATION D'INCLUSION

La structure hiérarchique de l'information repose sur la relation d'inclusion entre les individus statistiques.

Compte tenu de l'emboîtement des niveaux statistiques, Personne, Ménage, Parcelle : chaque personne (unité statistique c du niveau C) est incluse dans un ménage (unité statistique b du niveau B) qui, lui-même, est inclus dans un secteur (unité statistique a du niveau A):

$$\forall\, c \in C, \exists\, b \in B \text{ et } a \in A \text{ tel que } a \supset b \supset c$$

et sachant qu'une hiérarchie est donnée par une série de partitions emboîtées sur un ensemble fini I c'est-à-dire par une suite finie de partitions $\{P_0, P_1, \ldots, P_n\}$ d'autant moins fines que l'indice en est plus élevé ; une partition P_{n-1} d'un ensemble I est dite plus fine qu'une autre partition P_n du même ensemble si [1] :

$$\forall\, p' \in P_{n-1}, \exists\, p \in P_n : p \supset p'$$

alors **la relation d'inclusion entre les unités statistiques traduit la structure hiérarchique.**

Chaque parcelle représente une classe de ménages ou de personnes et chaque ménage constitue une classe de personnes. Par conséquent le niveau Parcelle est considéré comme une partition sur l'ensemble des ménages et sur celui des personnes ainsi que le niveau Ménage constitue une partition sur l'ensemble de personnes. Ceci nous amène à assimiler un niveau d'analyse à une partition et à considérer **une série de partitions emboîtées** engendrées par les populations des niveaux supérieurs sur les populations des niveaux plus fins.

La figure n°4 présente les diverses partitions dues à la structure hiérarchique entre les unités statistiques. Nous envisageons d'abord les partitions simples obtenues :

- sur l'ensemble B (les ménages) muni d'une partition engendrée par A (les parcelles). On note B_a, l'ensemble des unités statistiques b de B appartenant à une unité a de A et l'on a :

$$B = \cup \{B_a \mid a \in A\}$$

- sur l'ensemble C (les personnes) muni d'une partition engendrée par A (les parcelles). On note C_a, l'ensemble des unités statistiques c de C appartenant à une unité a de A et l'on a :

$$C = \cup \{C_a \mid a \in A\}$$

- sur l'ensemble C (les personnes) muni d'une partition engendrée par B (les ménages). On note C_b, l'ensemble des unités statistiques c de C appartenant à une même unité b de B et l'on a :

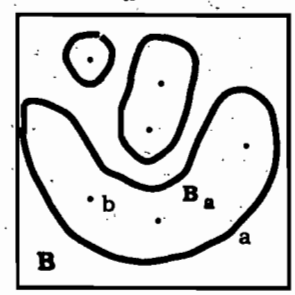
$$C = \cup \{C_b \mid b \in B\}$$

[1] [BENZECRI, 1973a] p131.

Les partitions simples

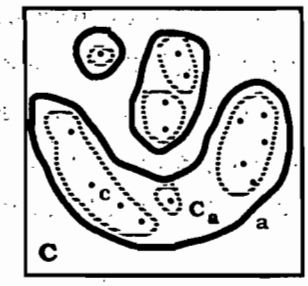
A= ensemble des parcelles, B= ensembles des ménages, C= ensemble des personnes

Sur l'ensemble B: $B = \cup \{B_a \mid a \in A\}$

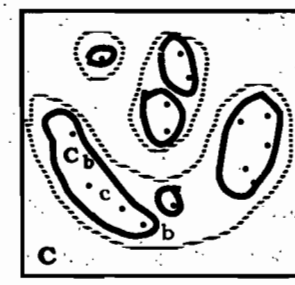


Sur l'ensemble C:

$C = \cup \{C_a \mid a \in A\}$

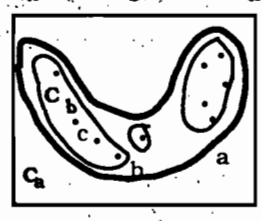


$C = \cup \{C_b \mid b \in B\}$



Les partitions emboîtées

$C_a = \cup \{C_b \mid b \in B_a\}$



$C = \cup \{\cup \{C_b \mid b \in B_a\} \mid a \in A\}$

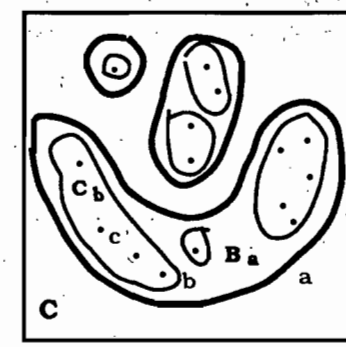


Figure n°4 : *Les différentes partitions issues de la structure hiérarchique de l'information*

Puis la série de partitions emboîtées est obtenue par l'ensemble A qui engendre une partition sur l'ensemble B qui lui-même engendre une partition sur l'ensemble C. On a:

d'une part : $C_a = \cup \{ C_b \mid b \in B_a \}$

et d'autre part : $C = \cup \{ \cup \{ C_b \mid b \in B_a \} \mid a \in A \}$

Par conséquent, il s'agit de travailler constamment sur des tableaux d'information dont l'ensemble des unités statistiques est partitionné.

2. 2. 3. LES RELATIONS D'AFFECTATION ET D'AGREGATION

Chaque niveau d'observation est décrit par un groupe de variables. Pour croiser ces groupes de variables entre eux, il faut les ramener à un même niveau d'analyse. Autrement dit, il faut établir **une relation entre le niveau de définition des variables et le niveau d'analyse.** Deux cas de figures se posent alors pour le transfert de l'information et nous identifions deux types de relations qui découlent de la relation d'inclusion entre les unités statistiques:

- **la "relation d'affectation"**[1] : lorsque le niveau d'analyse est plus fin que le niveau de définition des variables, il y a répétition ou affectation des informations des unités supérieures sur les unités plus fines (qui sont incluses dans les unités d'observation) : par exemple, le statut de résidence du ménage est affecté à chaque personne appartenant à ce ménage.

- **la relation d'agrégation** : lorsque le niveau d'analyse est moins fin que le niveau de définition des variables, il y a agrégation de l'information des unités plus fines vers le niveau moins fin (qui englobe les unités d'observation) : on somme l'information des personnes appartenant à un même ménage. Par exemple, on obtient une nouvelle variable au niveau Ménage qui est le nombre de personnes du ménage ayant une activité dans le secteur informel.

Afin de minimiser la perte d'information dans le transfert, il est naturel **d'examiner les phénomènes au niveau le plus fin et de "remonter" l'information vers les niveaux les moins fins.**

C'est sur ce principe que vont être construits les tableaux d'information qui appréhendent la structure hiérarchique de l'information. Pour cela ces relations vont nous permettre de mettre au point un

[1] [DUBOIS *et al*, 1989]

système de pondération cohérent pour généraliser un principe de construction de tableau en considérant les **deux choix à faire : l'un sur la population et l'autre sur le niveau d'analyse.**

2. 2. 4. UN SYSTEME DE PONDERATION

Travailler sur une **population homogène** à un niveau d'analyse donné, c'est faire le choix d'un **système de pondération**[1] c'est-à-dire d'un système de nombres strictement positifs sur l'ensemble des types d'unités.

Pour cela, on doit se fixer une **population statistique de référence**, c'est-à-dire choisir un niveau R (auquel correspond la population R) parmi les niveaux emboîtés A, B et C. On donne alors une pondération p_r à chaque unité de cette population : p_r=1 s'il n'y a aucun redressement à faire, dans quel cas $\sum_r p_r = n$ est le nombre d'unités statistiques composant la population d'étude.

Du fait de l'emboîtement des unités statistiques des différents types A, B et C, on doit avoir :

$$p_a = \sum_{b \in a} p_b \; ; \; p_b = \sum_{c \in b} p_c \text{ et l'on a naturellement } p_a = \sum_{c \in a} p_c = \sum_{b \in a} \sum_{c \in b} p_c \; .$$

Par conséquent,

$$\sum_r p_r = \sum_a p_a = \sum_b p_b = \sum_c p_c = n \; .$$

La figure n°5 présente un exemple de construction de tableaux d'information. Pour trois niveaux de collecte donnés, A, B, C, (respectivement Parcelle, Ménage, Personne par exemple), on a trois populations potentielles pour exploiter les variables. Une fois fixée la population de référence, R, on a au moins autant de tableaux qu'il y a de niveaux. On aboutit à la construction de 9 tableaux qui croisent l'ensemble des variables définies sur les trois niveaux ce qui élargit les possibilités d'analyse. On pose dans cet exemple $p_r = 1, \forall r \in R$ (on n'effectue aucun redressement). Le système de pondération est élaboré de la manière suivante.

Compte tenu de l'emboîtement des unités statistiques et de l'unité statistique de référence R, on en déduit toutes les pondérations à donner aux autres unités pour aboutir ainsi au système de pondération donné suivant le niveau de référence:

- si on travaille sur la population des unités c (les personnes), R = C, on pose $p_r = p_c$ et l'on a :

$$p_a = \sum_{c \in a} p_c \text{ et } p_b = \sum_{c \in b} p_c$$

[1] [BENZECRI, 1977b], p338.

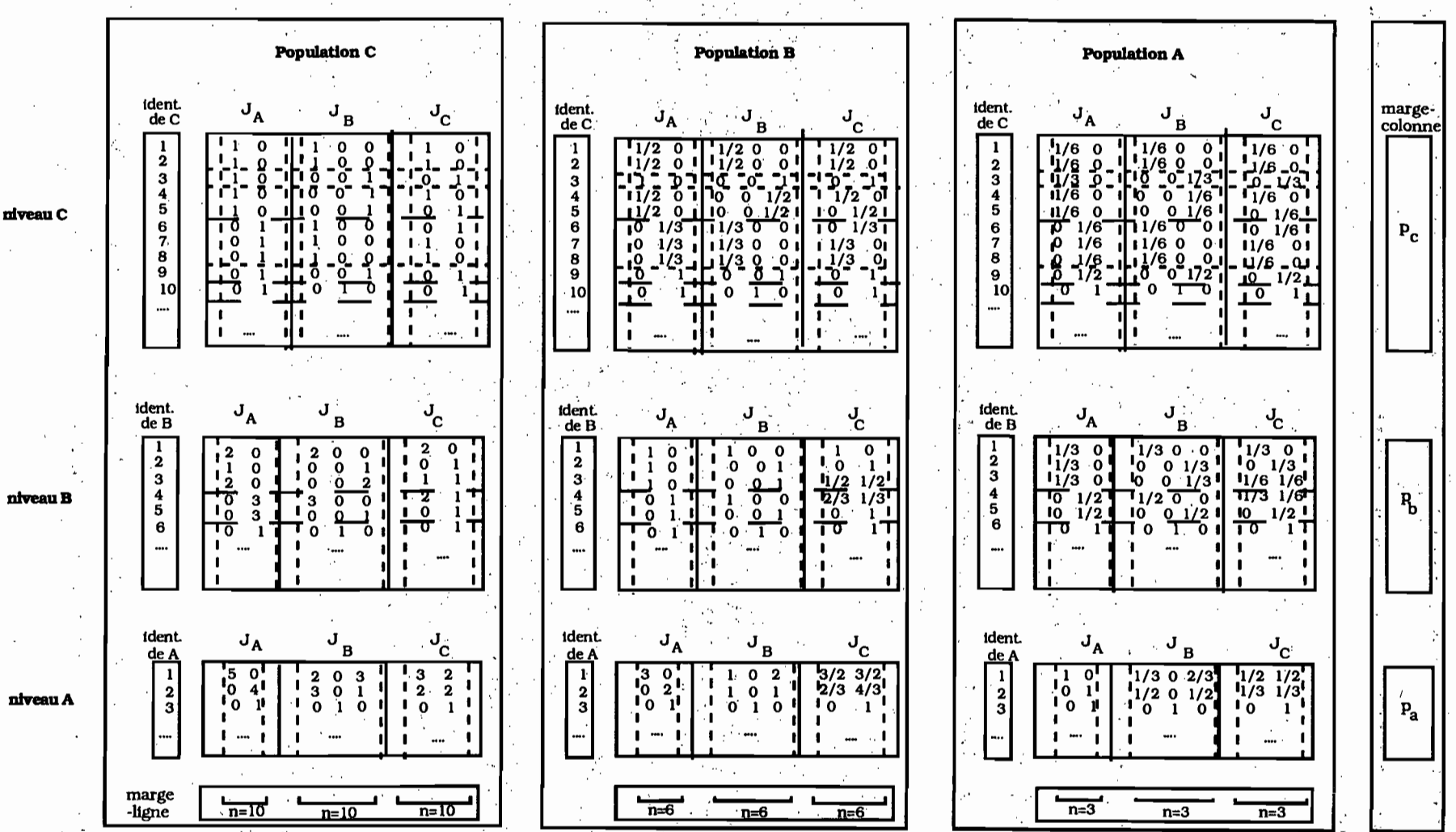


Figure n°5 : *Tableaux juxtaposant des variables qualitatives définies à plusieurs nivea ux*

par exemple si l'on pose $p_c = 1$ alors $\sum_c p_c = n$ (= 10) est le nombre de personnes, p_b est le nombre de personnes appartenant au ménage b et p_a est le nombre de personnes résidant sur la parcelle a.

- si on travaille sur la population des unités b (les ménages), R = B, on pose $p_r = p_b$ et l'on a :

$$p_a = \sum_{b \in a} p_b \quad \text{et} \quad p_c = \frac{p_b}{\sum_{c \in b} 1}$$

si l'on pose $p_b = 1$ alors $\sum_b p_b = n$ (= 6) est le nombre de ménages, p_c est le nombre de personnes pondéré par le nombre de personnes appartenant au ménage b et p_a est le nombre de ménages résidant sur la parcelle a.

- si on travaille sur la population des unités a (les parcelles), R = A, on pose $p_r = p_a$ et l'on a :

$$p_b = \frac{p_a}{\sum_{b \in a} 1} \quad \text{et} \quad p_c = \frac{p_b}{\sum_{c \in b} 1}$$

si l'on pose $p_a = 1$ alors $\sum_a p_a = n$ (= 3) est le nombre de parcelles, p_c est le nombre de personnes, appartenant au ménage b résidant sur la parcelle a, pondéré par le produit du nombre de personnes appartenant au ménage b et du nombre de ménages appartenant à la parcelle a. p_b est le nombre de ménages pondéré par le nombre de ménages résidant sur la parcelle a.

On vérifie à chaque fois que : $\sum_r p_r = \sum_a p_a = \sum_b p_b = \sum_c p_c = n$

On définit ainsi des systèmes de pondération qui ont la propriété :

- **d'homogénéiser la population statistique** : quelque soit le niveau d'analyse A, B ou C, on travaille sur la même population R;
- **d'établir des marges cohérentes** données par chaque élément du système de pondération : p_c (resp. p_b et p_a) correspond, à un coefficient près, aux marges en ligne du tableau associé au niveau d'analyse C (resp. B et A).

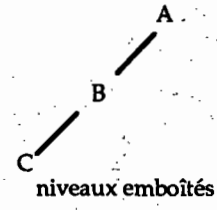
Remarque :

- Le choix de pondération pour une unité statistique, r, entraîne une pondération pour toutes les unités statistiques emboîtées de la hiérarchie.

En considérant ce système de pondération, on peut **exploiter simultanément un ensemble de variables définies à plusieurs niveaux suivant les différents niveaux d'analyse** dans la hiérarchie. Nous distinguons les deux types de niveaux ordonnés ou indépendants.

2. 2. 5. LES NIVEAUX EMBOITES

Une fois la population statistique fixée par un système de pondération, (p_a, p_b, p_c), on en déduit une série de tableaux qui permettent d'aborder les différents niveaux d'analyse emboîtés dans la hiérarchie.



On construit d'abord le tableau associé au niveau le plus fin, C, puis par cumul, on obtient les tableaux de correspondances associés aux niveaux supérieurs d'agrégation (cf figure n°5) :

- **au niveau le plus fin**, C, (la Personne), on construit **le tableau disjonctif complet** HC(C,J) **pondéré par p_c** et qui croise les unités statistiques de base, c (les personnes), avec l'ensemble des k variables éclatées en modalités, $J = J_A \cup J_B \cup J_C$:

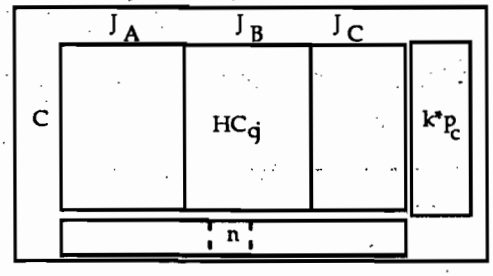


Figure n°6 : *Tableau disjonctif complet associé au niveau le plus fin, C*

où, $HC_{cj} = p_c$ ou 0 selon que l'unité statistique c possède ou non la modalité j de J.

Du fait que l'ensemble C est muni de deux partitions B et A, les informations des unités supérieures sont répétées ou réaffectées sur les unités inférieures qu'elles définissent et l'on a:

$$\forall a \in A ; \forall c \text{ et } c' \in a ; \forall j \in J_A : \qquad HC_{cj} = HC_{c'j}$$

$$\forall b \in B ; \forall c \text{ et } c' \in b ; \forall j \in J_A \cup J_B : \qquad HC_{cj} = HC_{c'j}$$

En tenant compte de cette série de partitions emboîtées A et B sur l'ensemble des unités statistiques c de C, on déduit ensuite, à partir du tableau disjonctif complet pondéré associé au niveau de base, les tableaux de correspondances (cf figure n°5) :

- **Au niveau intermédiaire**, B, (le Ménage), on crée le tableau de correspondances HB(B,J) qui croise les unités statistiques, b (les ménages), associées à ce niveau avec l'ensemble des modalités, J :

  $\forall j \in J$: $$HB_{bj} = \sum_{c \in b} HC_{cj}$$

  Ce tableau est obtenu en cumulant les informations des unités statistiques de base, c, appartenant à une même unité d'agrégation, b.

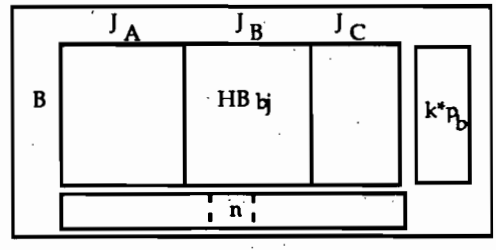


Figure n°7 : *Tableau des correspondances associé au niveau intermédiaire B*

- **Au niveau supérieur**, A, (la Parcelle), on a le tableau de correspondances HA(A,J) qui croise les unités statistiques, a (les parcelles), associées à ce niveau avec l'ensemble des modalités, J :

  $\forall j \in J$: $$HA_{aj} = \sum_{c \in a} HC_{cj} = \sum_{b \in a} \sum_{c \in b} HC_{cj} = \sum_{b \in a} HB_{bj}$$

  Ce tableau est obtenu soit par agrégation à partir du niveau de base, C, soit par réagrégation à partir du niveau intermédiaire, B.

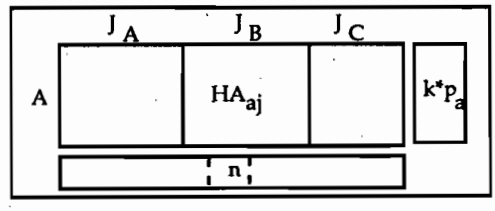


Figure n°8 : *Tableau des correspondances associé au niveau supérieur A*

On vérifie que les marges pour une même population R ont un sens :

- les marges en colonne : $HC_j = HB_j = HA_j$ et $\sum_{j \in q} HC_j = n$ où q est une variable
- les marges en ligne : $\sum_j HC_{cj} = k*p_c$ $\sum_j HB_{bj} = k*p_b$ $\sum_j HA_{aj} = k*p_a$

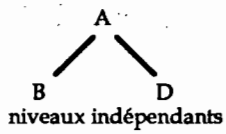
- l'effectif total : $k^*n = k^* \sum_r p_r = \sum_j HC_j = \sum_j HB_j = \sum_j HA_j$

Remarque :

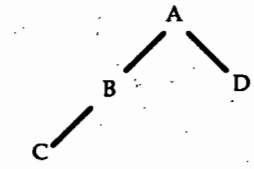
- Dans le cadre de niveaux géographiques, les tableaux de correspondances associés aux niveaux supérieurs sont appelés "matrices d'information spatiale".

2. 2. 6. NIVEAUX INDEPENDANTS ET GENERALISATION

Nous nous plaçons maintenant dans la situation où les niveaux B et D ou C et D sont indépendants et nous allons voir que ceci introduit une généralisation à la structure hiérarchique de l'information.



niveaux indépendants



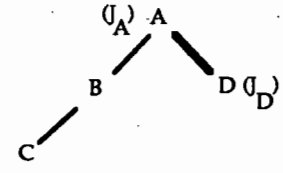
les différents types de niveaux, emboîtés et indépendants

Par exemple A (respectivement B, C et D) désigne le niveau Parcelle (respectivement Ménage, Personne et Bâtiment). La seule relation entre les niveaux B et D, ou C et D, est l'existence d'un niveau commun supérieur A qui les relie. **Le principe est alors d'agréger l'information contenue dans au moins un des niveaux indépendants à un niveau commun** afin d'homogénéiser la population statistique. Puis on se ramène à la situation de niveaux emboîtés. On procède en deux étapes :

1- On impose, pour au moins une des branches de l'arbre (par exemple A-D), l'unité statistique de référence qui correspond à l'unité supérieure a, commune aux deux chemins indépendants.

On pose alors : p_a et $p_d = \frac{p_a}{\sum_{d \in a} 1}$

On se trouve dans le cas de niveaux ordonnés. On agrège l'information du niveau le plus fin, D, au niveau supérieur commun, A. On augmente les variables de A, $J'_A = J_A \cup J_D$:



et on obtient les deux tableaux, HD(D,J) qui est un tableau disjonctif complet pondéré par p_d et HA(A,J) où $HA_{aj} = \sum_{d \in a} HD_{dj}$ avec $j \in J'_A$.

Par exemple pour 2 variables (k=2) et en posant $p_a = 1$:

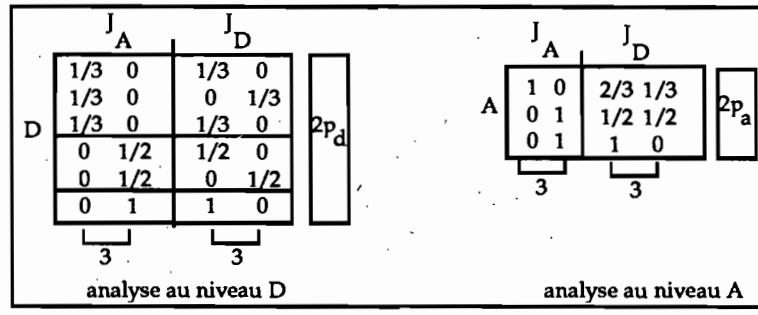
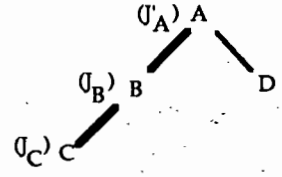


Figure n°9 : *Niveaux indépendants : 1ère étape*

on vérifie que le nombre d'unités a, définies par le niveau supérieur commun, est

$$\sum_d p_d = \sum_a p_a = n\ (=3).$$

2- Une fois ramené le niveau D au niveau commun A, on peut procéder comme pour des niveaux emboîtés. On fixe alors un système de pondération suivant la population statistique A, B ou C sur laquelle on désire travailler. On pose : $J = J'_A \cup J_B \cup J_C$



On obtient la série de tableaux de correspondances associés aux différents niveaux d'analyse :

- au niveau le plus fin, C, on a HC(CJ) où :
  $HC_{cj} = p_c$ ou 0 selon que l'unité statistique c possède ou non la modalité j de $J_C \cup J_B$ et
  $HC_{cj} = p_c\, HA_{aj}$ si $j \in J'_A$

- au niveau intermédiaire, B, on a HB(BJ) où $\forall j \in J : HB_{bj} = \sum_{c \in b} HC_{cj}$

- au niveau supérieur, A, on a HA(AJ) avec $\forall j \in J : HA_{aj} = \sum_{b \in a} HB_{bj}$

Par exemple pour 4 variables et en posant la population de référence associée au niveau C, p_c=1:

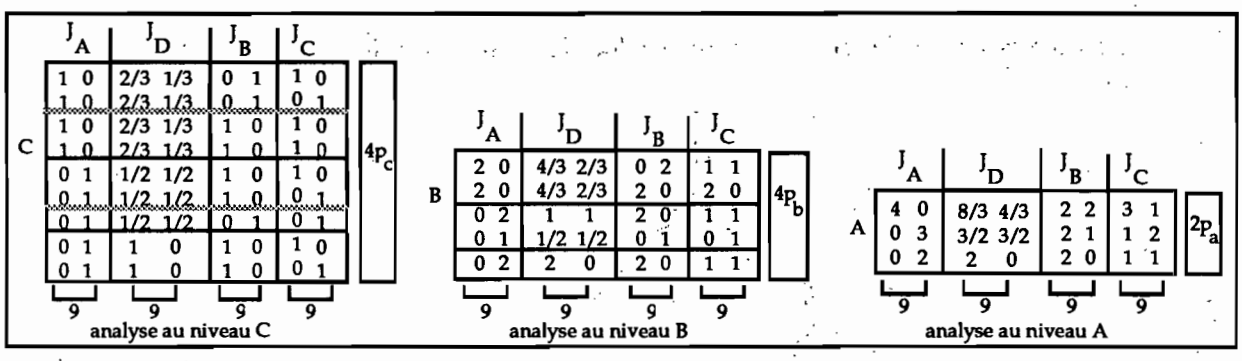


Figure n°10 : *Niveaux indépendants :2ème étape*

on vérifie que le nombre d'unités c de référence est $\sum_{c} p_c = \sum_{b} p_a = \sum_{a} p_a = n\ (=9)$.

Conclusion :

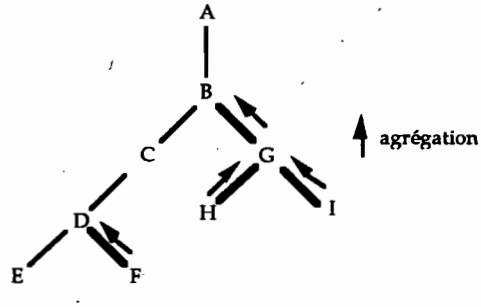
- L'étude de niveaux indépendants peut être facilement ramenée à une étude de niveaux emboîtés.

2. 2. 7. EXEMPLE D'UNE STRUCTURE COMPLEXE

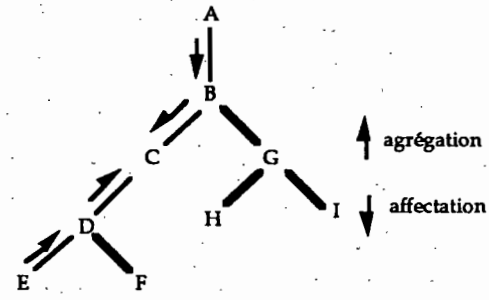
Nous présentons ici le principe pour appréhender une structure complexe. Les termes généraux des tableaux associés à cette structure ont été exposés plus longuement dans des travaux antérieurs [PIRON, 1990].

Soit un système d'information ayant 9 niveaux d'observation (A à I) et l'on a n le nombre de chemins dans l'arbre de la structure et *s* le chemin comprenant le niveau de référence, par exemple C :

- On considère d'abord les n-1 chemins correspondant aux niveaux indépendants et l'on impose comme unités de référence les unités supérieures communes qui appartiennent à *s*. A chaque chemin est associé un système de pondération.



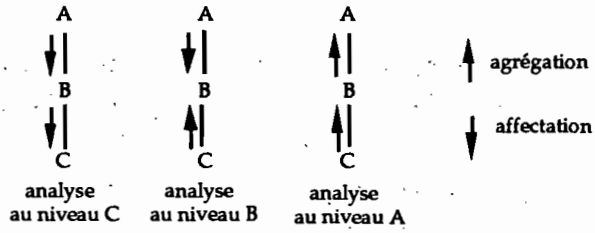
- On considère ensuite le chemin *s* et l'on se situe alors dans le cas de niveaux ordonnés.



2. 2. 8. SYNTHESE

A chaque niveau d'analyse correspond un tableau d'information. On aboutit à **une série de tableaux** qui juxtaposent ou **croisent l'ensemble des variables définies à plusieurs niveaux et qui traitent une même population statistique** donnée par un système de pondération cohérent. **Dans le cas de niveaux indépendants, on se ramène, par le processus d'agrégation, à des niveaux emboîtés.** Par conséquent on considère la relation d'inclusion entre les unités statistiques où l'on distingue:

- une **relation d'agrégation** qui permet de passer du niveau inférieur à un niveau supérieur;
- une **relation d'affectation** qui permet de passer du niveau supérieur à un niveau plus fin.



Ainsi, une fois fixée la population statistique, il en découle l'étude sur un ou sur l'ensemble des niveaux.

LES DIVERS TABLEAUX DU PROCESSUS D'AGREGATION

L'étude d'une structure hiérarchique de l'information aboutit à la construction d'une série de tableaux associés à chaque niveau d'analyse pour une population statistique homogène. Nous formalisons ainsi **le processus d'agrégation pour des variables qualitatives** ou considérées comme telle et nous allons pour cela identifier les différents types de tableaux. L'intérêt est d'obtenir, à partir de tableaux de données différents, des résultats identiques issus de l'analyse des correspondances.

2. 3. 1. FORMALISATION DU PROCESSUS D'AGREGATION

Soit l'ensemble C muni de deux partitions emboîtées, A et B, on considère à partir du tableau associé au niveau d'observation le plus fin, HC_{CJ}, ses correspondances HA_{AJ} et HB_{BJ} obtenues par :

$$HB_{bj} = \sum_{c \in b} HC_{cj}$$

$$HA_{aj} = \sum_{c \in a} HC_{cj}$$

mais on a également :

$$HA_{aj} = \sum_{b \in a} HB_{bj} = \sum_{b \in a} \sum_{c \in b} HC_{cj}$$

Cela signifie qu'un niveau d'agrégation est déduit à partir du niveau d'observation de base ou d'un niveau d'agrégation plus fin. L'information des unités statistiques plus fines appartenant à une même classe est cumulée. De ce fait, les unités du niveau inférieur sont assimilées et donc analysées au travers du centre de gravité de leur classe d'appartenance. Agréger rend compte d'une valeur moyenne de l'information par classe dont l'ensemble définit le niveau d'agrégation.

Nous allons présenter les différents tableaux possibles qui traitent des données agrégées et leur interprétation. Pour cela, nous allons regarder les conséquences de la structure hiérarchique de l'information sur les tableaux disjonctifs complets, pondérés ou non, qui correspondent au niveau d'observation le plus fin.

2. 3. 2. PRESENTATION DES TABLEAUX ASSOCIES AUX DIFFERENTS NIVEAUX D'ANALYSE

Pour chaque niveau d'analyse, A ou B, définissant une partition sur les niveaux inférieurs, B ou C, on associe trois types de tableaux. De dimensions différentes, ils appréhendent la même population R donnée par un système de pondération pour le même ensemble de variables J. Par conséquent c'est l'espace des individus(-lignes) I qui varie suivant la mesure utilisée. Nous verrons dans la troisième partie, une méthode pour l'analyse d'un système d'échelles, l'intérêt de disposer de tels tableaux de dimensions différentes sur lesquels effectuer une analyse factorielle des correspondances aboutit, compte tenu du principe de l'équivalence distributionnelle[1], aux mêmes résultats c'est-à-dire aux mêmes facteurs.

2. 3. 2. 1. Conséquences de la structure hiérarchique sur les tableaux disjonctifs complets

Dans un tableau disjonctif complet, chaque variable, par ses indicatrices ou modalités, définit une partition sur l'ensemble des individus statistiques ; par exemple la variable "secteur d'activité" éclatée en trois modalités, détermine trois classes de personnes : celles travaillant dans le secteur informel, dans le secteur moderne et celles sans profession.

Par ailleurs, chaque unité statistique est affectée d'un identificateur appelé également "variable de filiation"[2] qui permet de définir le lien hiérarchique entre les unités de niveaux différents ; chaque personne est reliée à un ménage qui lui-même est relié à une parcelle par le numéro du ménage puis par celui de la parcelle, chacun déterminant respectivement les identificateurs du ménage et de la parcelle.

L'identificateur, q_A (resp. q_B), affecté aux unités définissant le niveau supérieur, A (resp. B), joue donc le rôle de la variable de partition sur les populations associées aux niveaux inférieurs, B et C (resp.C).

Par conséquent, chaque modalité a de la variable q_A (resp. q_B) est assimilée à une unité statistique de A (resp.B) et constitue une classe associée au niveau supérieur A (resp. B).

---

[1][BENZECRI, 1973a]
[2][DUBOIS *et al*, 1989]

2. 3. 2. 2. Tableaux de correspondances ou bandes du tableau de Burt

Les partitions dues à l'emboîtement des niveaux statistiques se traduisent sur le tableau disjonctif complet HC_{CJ} par les tableaux des indicatrices, HC_{CA} et HC_{CB} associées respectivement aux modalités des variables q_A et q_B (cf figure n°11).

On construit par cumul à partir de HC_{CJ} les tableaux de correspondances HA_{AJ} et HB_{BJ}. Ce sont des tableaux du type (unités, variables) où les unités a (resp. b) sont des unités d'analyse pas forcément observées et déjà agrégées : elles sont constituées par le cumul de l'information des unités statistiques plus fines, c, leur appartenant et définissent ainsi des classes, C_a (resp. C_b). Il s'agit alors des bandes du tableau de Burt $BC_{J'J'}$ associé au tableau disjonctif complet $HC_{CJ'}$ où $J' = J \cup q_A \cup q_B$ qui croisent les variables avec les variables de filiation q_A et q_B(cf figure n°12). Elles sont notées BC_{AJ} et BC_{BJ}. En effet si on pose $p_r = p_c = 1$:

$\forall b \in J_B, \forall j \in J : HB_{bj} = \sum_{c \in b} HC_{cj} = \sum_{C} \frac{HC_{cj}\, HC_{cb}}{HC_c} = BC_{bj}$. C'est le nombre d'unités statistiques c contenues dans b ayant adopté la modalité j de J ;

$\forall a \in J_A, \forall j \in J : HA_{aj} = BC_{aj} = \sum_{C} \frac{HC_{cj}\, HC_{ca}}{HC_c} = \sum_{c \in a} HC_{cj}$. C'est le nombre d'unités statistiques c contenues dans a ayant adopté la modalité j de J.

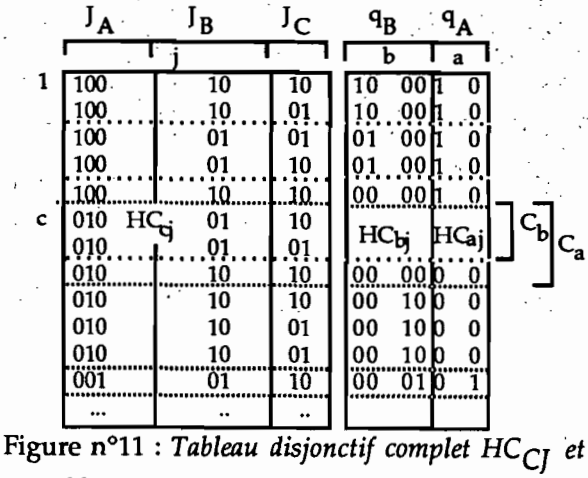


Figure n°11 : *Tableau disjonctif complet HC_{CJ} et tableaux des indicatrices, HC_{BJ} et HC_{AJ}*

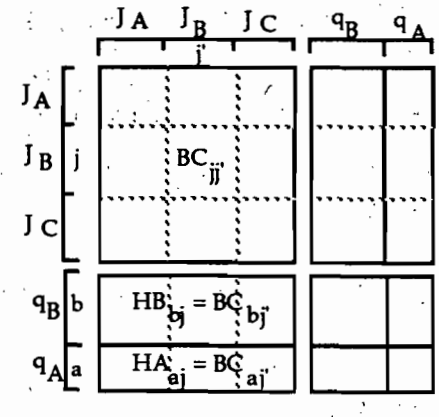


Figure n°12 : *Tableau de Burt, BC_{JJ} et tableaux des correspondances, HB_{BJ} et HA_{AJ}*

On peut également construire HA_{AJ} à partir de HB_{BJ} de la manière suivante :

$\forall a \in J_A, \forall j \in J : HA_{aj} = \sum_{B} \frac{HB_{bj}\, HB_{ba}}{HB_b} = \sum_{b \in a} HB_{bj}$; c'est le nombre d'unités statistiques c, contenues dans les unités b elles-mêmes contenues dans les unités a, ayant adopté la modalité j de J.

Par conséquent :

- **au niveau le plus fin, C**, est associé le tableau disjonctif complet HC_{CJ} **défini sur C*J.**
- **au niveau intermédiaire, B** est associé le tableau de correspondances HB_{BJ} c'est-à-dire la bande du tableau de Burt BC_{BJ}, **défini sur le produit B*J.**
- **au niveau supérieur, A** est associé le tableau de correspondances HA_{AJ} c'est-à-dire la bande du tableau de Burt BC_{AJ}, **défini sur le produit A*J.**

Par exemple au niveau Parcelle on a :

		J_A	J_B	J_C
		j		
A	1	500	32	32
		030	12	21
	a	030	HA_{aj} 30	12
		001	01	10
		...	...	...

où $HA_{aj} = BC_{aj} = \sum_{c \in a} HC_{cj}$

L'interprétation de ces tableaux est immédiate puisqu'il s'agit de ventiler une population statistique suivant les deux critères donnés par les unités statistiques ou encore une classe définissant le niveau d'agrégation et par la variable j.

2. 3. 2. 3. Tableau des centres de gravité

La bande du tableau de Burt, croisant la variable de filiation correspondant à un niveau d'agrégation donné, A ou B, avec l'ensemble J des variables, est le tableau des centres de gravité des classes associées à ce niveau. Ce tableau est défini sur A*J ou sur B*J. On construit également une autre **tableau des centres de gravité des classes** mais cette fois **défini sur C*J** quelque soit le niveau d'analyse considéré :

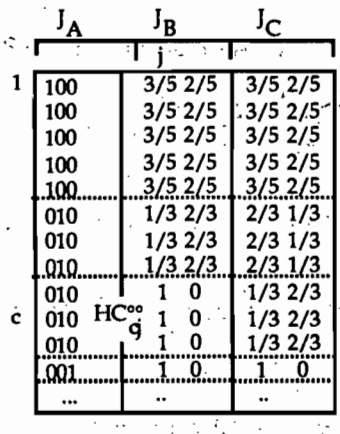
- au niveau intermédiaire B, le tableau des centres de gravité correspondant, HC°, a pour terme général:

  $\forall j \in J$ et $\forall c \in b$: $\quad HC^{\circ}_{cj} = \dfrac{HB_{bj}}{p_b}$

- au niveau supérieur A, le terme général du tableau des centres de gravité correspondant, HC°°, est donné par :

  $\forall j \in J$ et $\forall c \in a$: $\quad HC^{\circ\circ}_{cj} = \dfrac{HA_{aj}}{p_a}$

  par exemple au niveau Parcelle on a :



Chaque unité statistique c (une personne) est représentée par le profil moyen de la classe a (la parcelle) à laquelle elle appartient.

Ces tableaux appréhendent la même population, ils sont de même dimension c'est-à-dire définis sur le même produit C*J mais ont une mesure différente exprimant les centres de gravité associés au niveau d'analyse considéré.

Le processus d'agrégation revient à analyser les unités statistiques associées au niveau le plus fin, C, au travers de leur classe d'appartenance définissant le niveau d'agrégation.

2. 3. 2. 4. Tableau de "Burt"

D'une manière générale, un tableau de Burt est obtenu à partir du tableau disjonctif complet. Mais on peut tout à fait concevoir de construire un tableau assimilé à un tableau de Burt à partir d'un tableau de correspondances par la même formule : $B_{jj'} = \sum_i \frac{K_{ij} K_{ij'}}{K_i}$

Remarque :

- Dans le cas où K_{IJ} est un tableau disjonctif complet, les unités i désignent les unités d'observation associées au niveau le plus fin ne pouvant être décomposées. Dans le cas où K_{IJ} est un tableau de correspondances, ces unités i sont des unités agrégées définissant des classes.

- Si la population de référence est définie par le niveau d'observation le plus fin, alors les unités i du tableau disjonctif complet correspondent aux individus statistiques composant la population d'étude. En revanche, les unités d'analyse du tableau de correspondances ne sont pas assimilées aux individus statistiques qui composent la population d'étude.

Si l'on considère les trois niveaux d'observation emboîtés A, B, C alors :

- au niveau le plus fin, C, on crée le tableau de burt, noté BC_{JJ} à partir du tableau disjonctif complet HC_{CJ} : $BC_{jj'} = \sum_{c} \frac{HC_{cj}\, HC_{cj'}}{HC_c}$
- au niveau intermédiaire B, on construit le tableau assimilé à un tableau de Burt, BB_{JJ}, à partir du tableau de correspondances HB_{BJ} par la formule : $BB_{jj'} = \sum_{b} \frac{HB_{bj}\, HB_{bj'}}{HB_b}$
- au niveau supérieur A, on construit le tableau assimilé à un tableau de Burt, BA_{JJ}, à partir du tableau de correspondances HA_{AJ} par la formule : $BA_{jj'} = \sum_{a} \frac{HA_{aj}\, HA_{aj'}}{HA_a}$

Il s'agit alors d'interpréter les données issues de tels tableaux (cf annexe 2. 1).

2. 3. 2. 5. Synthèse

Tous les types de tableaux proposés respectent les conditions imposées par l'analyse des correspondances : tableau rectangulaire de données homogènes dont les marges sont cohérentes et sur lequel peut être réalisée une analyse factorielle des correspondances afin de conserver ses propriétés (distance du chi-deux, équivalence distributionnelle ...).

Compte tenu de cette dernière propriété, l'équivalence distributionnelle, les analyses factorielles des correspondances appliquées aux trois types de tableaux associés à un même niveau d'analyse sont équivalentes. La figure n°11 représente ces tableaux qui traitent de données qualitatives et agrégées.

2.3.3. EXEMPLE D'INTERPRETATION

Afin de mettre en évidence l'interprétation des tableaux de Burt ou assimilés en tant que tel, nous présentons un exemple sur les deux niveaux emboîtés Parcelle-Ménage. On considère d'une part une variable définie au niveau Parcelle, l'électricité éclatée en deux modalités suivant si elle est installée ou non sur la parcelle et d'autre part une variable définie au niveau Ménage, le statut d'occupation ayant trois modalités. Nous envisageons les deux populations statistiques sur lesquelles il est possible de travailler :

- la population des Ménages,
- la population des Parcelles.

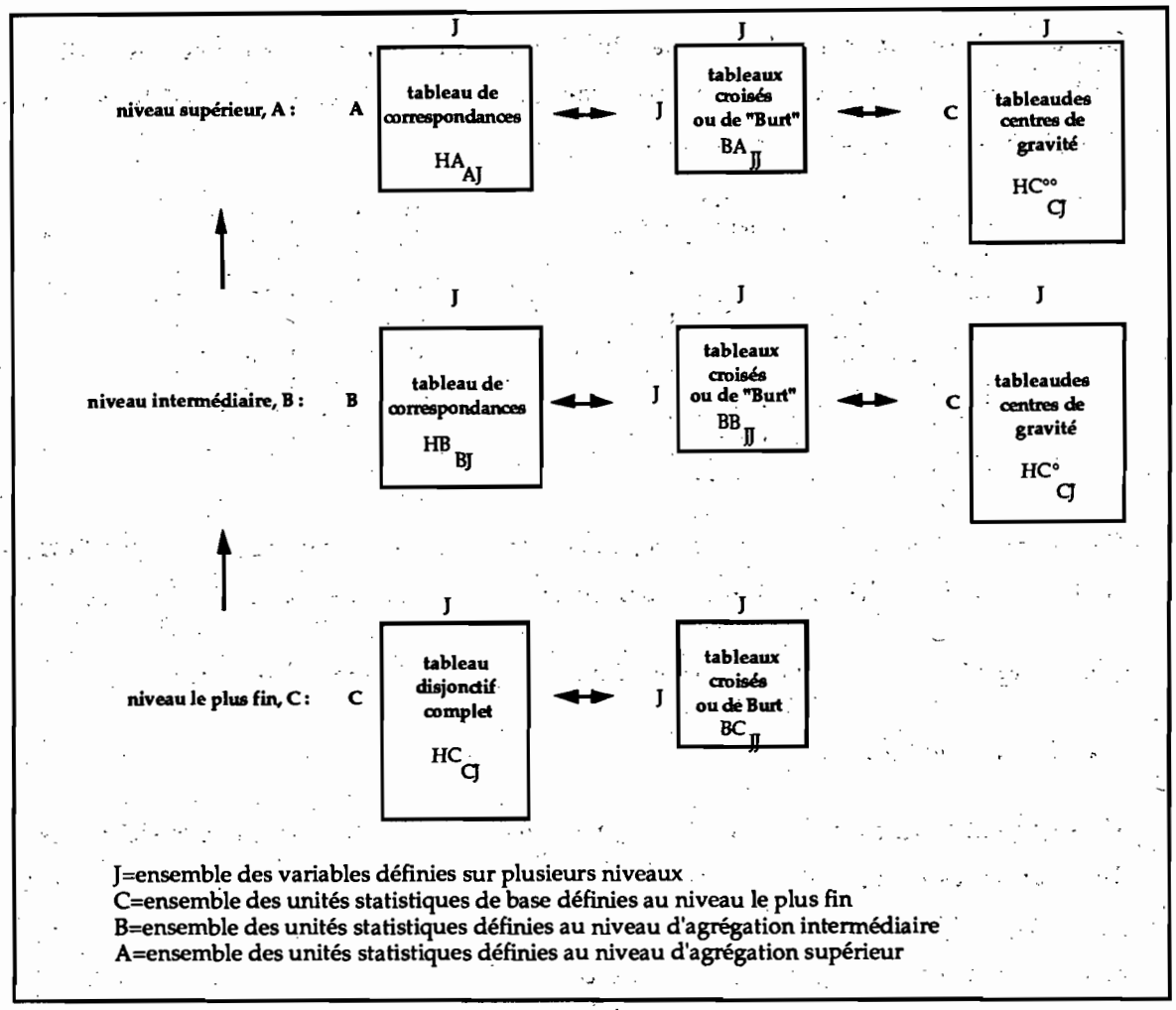


Figure n°13 : *Le processus d'agrégation pour des données qualitatives*

Ceci élargit les possibilités d'analyse puisqu'en général on étudie la population correspondant au niveau le plus fin, ici le Ménage, compte tenu des variables considérées.

Une fois fixée la population de référence, nous avons autant de tableaux qu'il y a de niveaux d'analyse. Dans cet exemple, nous considérons comme niveaux d'analyse les deux niveaux d'observation considérés.

2. 3. 3. 1. La population des Ménages :

L'unité statistique de référence est le Ménage. On comptabilise le nombre de ménages et l'on a le système de pondération : $p_m = 1$ et $p_p = \sum_{m \in p} 1$.

- **A l'échelle du Ménage :**

On construit un tableau à l'échelle la plus fine compte tenu des variables c'est-à-dire au niveau Ménage. On ne perd pas d'information. Les variables Parcelle sont répétées autant de fois qu'il y a de ménages sur la parcelle.

n°mén.	var. Parcelle elect	non el	...	var. Ménage prop	loc	héb	...	n°parc.
1	1	0		1	0	0		1
2	1	0		0	0	1		1
3	1	0		0	0	1		1
4	1	0		0	0	1		2
5	0	1		0	1	0		3
6	1	0		1	0	0		3
7	1	0		0	1	0		3
8	0	1		0	0	1		3
9	0	1		0	0	1		4
10	0	1		0	0	1		4

		var. Parcelle elect	var. Parcelle non el	var. Ménage pro	var. Ménage loc	var. Ménage héb
var. Parcelle	elect	6	0	2	1	3
	non el	0	4	0	1	3
var. Ménage	pro			2	0	0
	loc			0	2	0
	héb			0	0	6

A l'examen du tableau croisé, on obtient, au croisement des deux modalités "elect" et "heb", le nombre de ménages hébergés qui ont l'électricité sur la parcelle (c'est-à-dire 3 ménages).

C'est ce type de tableau que l'on exploite pour le traitement de données d'enquête.

- **A l'échelle de la Parcelle :**

Mais on peut aussi concevoir une exploitation des variables à l'échelle de la Parcelle, c'est-à-dire à un niveau d'agrégation. C'est le cas dès que l'on veut cartographier des résultats. On agrège alors l'information et on recrée un tableau (Parcelles, Variables) où les variables décrivent **simultanément** les ménages et les parcelle.

n°parc.	var. Parcelle elect	non el	...	var. Ménage prop	loc	héb	...
1	4	0		1	0	3	
2	0	1		0	1	0	
3	2	0		1	1	0	
4	0	3		0	0	3	

		var. Parcelle elect	var. Parcelle non el	var. Ménage pro	var. Ménage loc	var. Ménage héb
var. Parcelle	elect	6	0	2	1	3
	non el	0	4	0	1	3
var. Ménage	pro			0,8	0,5	0,7
	loc			0,5	1,5	0
	héb			0,7	0	5,3

Au croisement des modalités "elect" et "heb", on a comme précédemment le nombre de ménages hébergés qui ont l'électricité sur la parcelle. Mais au croisement des modalités "prop" et "heb", on a le nombre moyen (0,7) de ménages propriétaires et hébergés sachant qu'ils résident sur une même parcelle. Il s'agit d'une moyenne conditionnelle.

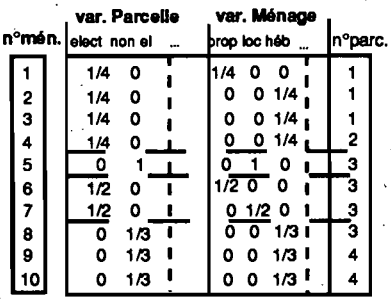
C'est ce type de tableau c'est-à-dire associé à un niveau d'agrégation que l'on exploite dès que l'on veut cartographier des résultats.

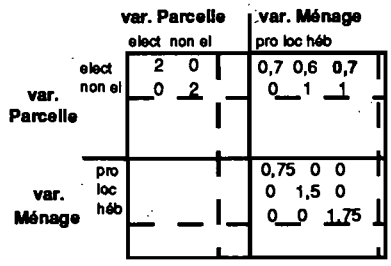
2. 3. 3. 2. La population des Parcelles :

On envisage, maintenant l'étude à partir de la population des parcelles qui devient la population de référence. On compte ici des parcelles. Le système de pondération est le suivant : $p_p=1$ et $p_m= \frac{1}{\sum_{m \in p} 1}$

Les tableaux sont construits comme précédemment et l'on obtient :

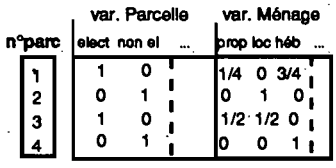
- **A l'échelle du Ménage :**

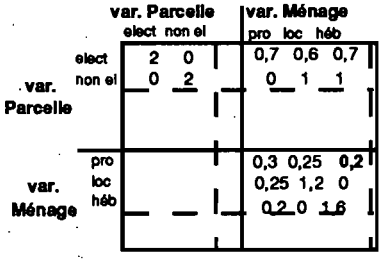




Au croisement des deux modalités "elect" et "heb", on a le nombre (0,7) de parcelles qui ont l'électricité et qui abritent des hébergés.

- **A l'échelle de la Parcelle :**





Au croisement des modalités "prop" et "heb", on compte le nombre moyen (0,2) de parcelles qui abritent en même temps des propriétaires et des hébergés.

VERS L'ANALYSE D'UN SYSTEME D'ECHELLES

La mise en évidence d'un système de pondération aboutit à une série de tableaux qui permettent d'exploiter simultanément un ensemble de variables définies à plusieurs niveaux suivant les différents niveaux d'analyse et qui traitent une même population statistique. Ce système repose sur la relation d'inclusion entre les niveaux statistiques, laquelle induit deux types de relations entre le niveau de définition des variables et le niveau d'analyse, la relation d'agrégation et celle d'affectation.

A partir de ces tableaux, on peut imaginer tous les croisements possibles entre les groupes de variables pour une population statistique et un niveau d'analyse donnés. De plus, dans la mesure où l'étude de niveaux indépendants peut facilement être ramenée à l'étude de niveaux emboîtés, **l'analyse d'un système d'information peut être envisagée par celle d'un système hiérarchique de l'information.**

Décrire une structure hiérarchique de l'information revient alors à considérer un système d'échelles comportant un nombre variable de niveaux.

L'information est agrégée : elle est perçue et donc mesurée différemment. Aussi, ce n'est pas un même phénomène qui est étudié sur plusieurs échelles mais ce sont des phénomènes différents car appréhendés à des échelles différentes ; ceci, aussi bien pour un groupe de variables spécifiques d'une population que pour le croisement de plusieurs groupes de variables, dans quel cas, ils subissent obligatoirement un changement d'échelle.

Par conséquent, **l'analyse de données agrégées** pose le problème de la pertinence des résultats si l'on ne mesure pas **l'effet du à l'agrégation,** c'est-à-dire si l'on n'étudie pas **la variabilité interne des classes** constituant le niveau d'agrégation. C'est ce que nous nous proposons d'étudier maintenant.

Troisième Partie

UNE METHODE POUR L'ANALYSE D'UN SYSTEME D'ECHELLES

L'analyse d'un système d'échelles doit permettre non seulement de dégager les structures spécifiques à chaque niveau d'analyse mais aussi de mettre en évidence les transformations de ces structures lorsque l'on passe d'un niveau à un autre et les pertes d'information associées à la seule prise en compte d'une échelle gobale.

Nous définissons un système d'échelles comme un ensemble de niveaux d'observation entre lesquels il existe une relation d'inclusion : c'est un système hiérarchique de l'information. Nous rappelons alors que pour étudier deux ou plusieurs niveaux indépendants, il faut les ramener à un niveau supérieur commun et l'on se place ainsi dans la situation de niveaux emboîtés, c'est-à-dire d'une relation d'inclusion entre les niveaux.

La démarche de l'étude d'un système d'échelles suppose que les classes définissant le niveau d'agrégation soient déjà établies. Par conséquent, on ne cherche donc pas, comme pour une classification, à construire des classes ne comprenant que des individus semblables entre eux[2], suivant des critères d'agrégation restant à définir (pour un tableau de correspondances, le critère d'agrégation maximise l'inertie inter-classes). Ici, ce critère est imposé par le lien hiérarchique et **une classe constitue alors un agrégat d'individus plus ou moins homogène.**

Le processus d'agrégation transforme l'information puisque celle-ci est analysée non plus suivant l'unité élémentaire d'observation dont elle est issue mais suivant une classe à laquelle elle appartient et qui définit le niveau d'agrégation. Il est alors important de saisir la nature de la transformation de l'information et ce quelque soit le découpage arbitraire ou non de l'univers d'étude. En effet :

- soit le découpage est raisonné, par exemple pour la constitution de groupes sociologiques, et il est intéressant de connaître la variabilité interne des différents groupes,
- soit le découpage est arbitraire, ce qui est souvent le cas en géographie, et cela pose le problème de l'hétérogénéité interne des unités spatiales sur lesquelles est fondée l'information géographique.

Par conséquent et par analogie à la présentation de R. CHAPUIS d'un système socio-spatial[1] il s'agit de comprendre, au travers d'un système hiérarchique de l'information, comment des individus statistiques (des ménages par exemple) agissent et interagissent dans des groupes (les parcelles) et comment ces groupes agissent et interagissent sur des groupes moins fins (les secteurs), pour ne citer que trois niveaux d'interactions.

[1][JAMBU, 1978]

[1][25ème Congrès de l'Union Géographique Internationale, août 1984]

Aussi, avec l'idée du changement d'échelle, il est tout aussi intéressant de pouvoir **dégager les structures spécifiques de chaque niveau, que celles que l'on perd dans le changement d'échelle,** même si les structures sont sensiblement équivalentes d'un niveau à un autre du fait d'une homogénéité éventuelle des groupes. C'est ce qui contribue à mettre en évidence toute **l'organisation du système d'information.** Nous nous attachons alors à regarder simultanément les étapes d'un système d'échelles afin de mieux cerner son organisation et permettre une lecture intégrée des différents facteurs aux différentes échelles. A partir de ce constat, on se demande :

- Quelle est l'influence d'une échelle sur la perception d'un phénomène donné, autrement dit quel niveau d'analyse choisir et quelles sont les conséquences de ce choix ?
- A quel niveau une variable perd-elle son pouvoir de discrimination ? Quelle information perd-on dans le passage d'un niveau à un autre ?
- Trouve-t-on les mêmes éléments d'organisation aux différentes échelles ?

A ce stade de la réflexion, il est important de préciser que le choix de l'échelle n'est pas neutre. Pour deux systèmes d'échelles différents mais de découpages identiques sur un même ensemble, on peut avoir des comportements différents : les classes associées à un même niveau sont globalement plus ou moins homogènes et une caractéristique peut intervenir pour un seul des niveaux. A.S.FOTHERINGHAM et D.W.S.WONG[1], au travers des modèles de régression linéaire multiples, observent que les méthodes multivariées ne sont pas fiables, d'un point de vue modélisateur, pour traiter du problème de l'unité zonale modifiable. A cet effet, ils ont montré que les résultats étaient très variables suivant les échelles et suivant des découpages différents obtenus pour une même échelle; d'où la mise en garde pour les études réalisées sur un seul niveau d'agrégation où l'on n'a pas de référence pour estimer les homogénéités internes et externes des unités spatiales.

- Peut-on alors définir, par les techniques descriptives d'analyses multivariées, un critère d'homogénéité à un niveau donné?

Toutes ces questions et ces remarques nous amènent à aller plus loin dans l'étude du changement d'échelle. **L'étude de la décomposition de l'inertie pour une série de partitions emboîtées** doit permettre de décrire les phénomènes spécifiques non seulement à un niveau mais aussi entre deux niveaux et d'appréhender simultanément plusieurs niveaux d'observation emboîtés c'est-à-dire un système d'échelles. Nous cherchons à étudier à la fois **les différenciations entre et à l'intérieur des classes constituant le niveau d'agrégation.**

[1][FOTHERINGHAM, WONG, 1991]

LES PRINCIPES DE LA DECOMPOSITION DE L'INERTIE

3. 1. 1. CONDITIONS SUR LE NUAGE DE POINTS POUR UNE DECOMPOSITION

Un nuage de points ne peut se décomposer que s'il existe une partition sur au moins un des deux ensembles I et J. Rappelons les deux principales structures de tableaux de correspondances pouvant être soumises à une décomposition de l'inertie et sur lesquelles peut être effectuée une analyse factorielle des correspondances, ceci afin de conserver ses propriétés :

- **Les tableaux ternaires** sont les plus courants et ont fait l'objet de nombreuses études[1]. Ils sont construits suivant un couple d'ensembles (I,J) sur lequel intervient un troisième ensemble T. Il s'agit donc d'une suite de tableaux binaires de même dimension indicés par T (souvent exprimant le temps) ou de tableaux de contingence définis par les mêmes caractères I et J sur des populations de types différents déterminés par T. Ainsi, T engendre une partition sur un des ensembles I ou J. Nous insistons sur le fait que ce type de tableau est, par conséquent, décomposé en sous-tableaux de même dimension : les classes indicées par la variable T sont composées des mêmes éléments d'ensemble, I et J (même produit I*J). L'ensemble I est partitionné suivant une variable T ∈ Q donnée. La structure initiale du tableau prend la configuration suivante[2]:

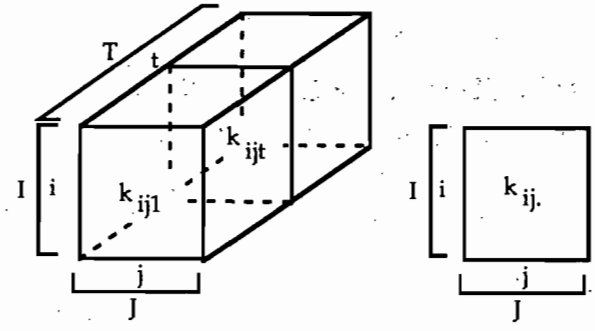


Figure n°14 : *Tableau ternaire et tableau somme sur T*

[1][BENZECRI, 1983], [ESCOFIER, 1982], [ESCOFIER, 1983]
[2][ESCOFIER, 1988] p168.

- **Les tableaux disjonctifs complets** pour lesquels, nous avons vu, une décomposition de l'inertie était tout à fait envisageable dans la mesure où la population d'un tel tableau est constamment partitionnée par les variables elles-mêmes (chaque variable, T, dans un tableau disjonctif complet définit, par ses modalités ou indicatrices, une partition sur les individus).

  Ce tableau se structure de la manière suivante où $I = \cup \{ I_t \mid t \in T \}$:

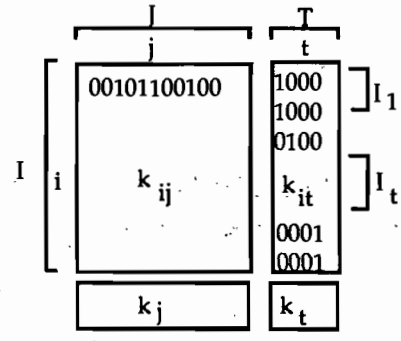


Figure n°15 : *Partition sur un tableau disjonctif complet*

Nous rappelons que la structure hiérarchique de l'information induit une série de partitions emboîtées engendrées par les populations des niveaux supérieurs sur les populations associées aux niveaux plus fins et qu'une partition est assimilée au niveau d'agrégation, une classe de la partition correspondant alors à une unité statistique associée à ce niveau..

Remarque :

- Dans cette dernière situation (tableau disjonctif complet), les classes d'une partition induite par une variable T définissent des sous-tableaux $K_{I_t J}$ de dimensions différentes. De plus, il se peut qu'une classe ne soit définie que par un seul élément. Ceci n'est plus vrai pour les tableaux ternaires.

3. 1. 2. DECOMPOSITION DE L'INERTIE SUIVANT UN NIVEAU D'AGREGATION

3. 1. 2. 1. Le principe

La décomposition de l'inertie d'un nuage de points structuré en sous-nuages (classes) rend compte des structures internes ou externes des classes d'une partition. Le principe de Huygens démontre que l'inertie totale est la somme de l'inertie inter-classes (inertie des centres de gravité des sous-nuages) et des inerties intra-classes (somme des inerties de chaque sous-nuage autour de leur centre de gravité).

$$\text{Inertie totale} = \text{Inertie inter} + \sum \text{Inertie intra}$$

L'inertie inter-classes mesure la variabilité entre les classes et l'inertie intra-classes quantifie la variabilité interne des classe (cf annexe 2.2).

3. 1. 2. 2. Exemple

Si on étudie l'ensemble de ménages suivant le niveau d'agrégation, le Secteur, schématisé par la figure suivante :

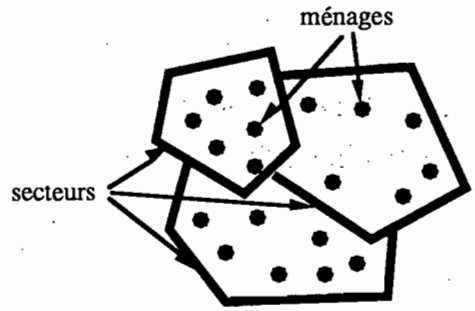


Figure n°16 : *Emboîtement Ménage-Secteur*

alors l'inertie du nuage des ménages se décompose de la manière suivante :

inertie de l'ensemble des ménages = inertie inter-secteurs + inertie intra-secteurs

où

- l'inertie inter-secteurs exprime **la variabilité entre les secteurs,** c'est-à-dire l'écart des secteurs par rapport au centre de gravité de l'ensemble des secteurs (qui est aussi celui de l'ensemble de ménages);
- l'inertie intra-secteurs mesure **la variabilité des ménages à l'intérieur des secteurs,** c'est-à-dire l'écart des ménages par rapport à leur secteur de résidence (centre de gravité de tous les ménages appartenant à un même secteur).

Nous pouvons représenter cette décomposition de la manière suivante :

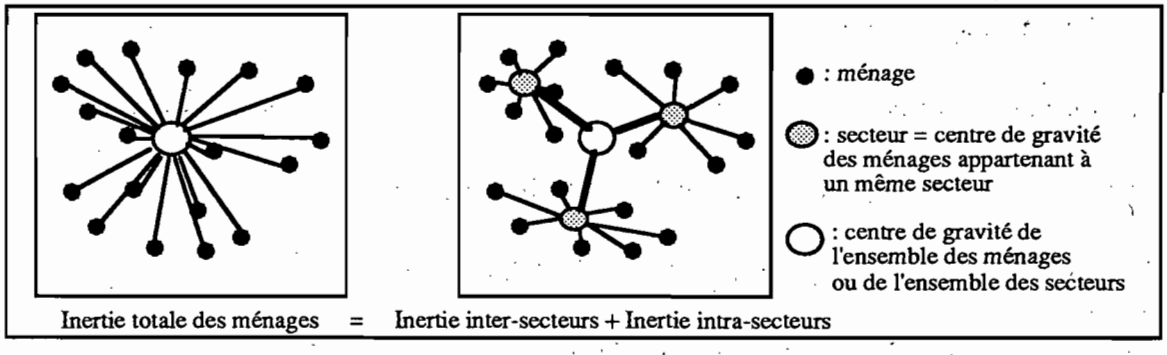


Figure n°17 : *Décomposition de l'inertie suivant le niveau d'agrégation, le Secteur*

Si les secteurs sont homogènes c'est-à-dire si les ménages au sein d'un même secteur ont un comportement identique quant aux variables étudiées, l'inertie inter-secteurs est élevée et par conséquent l'inertie intra-secteurs est faible indiquant ainsi la variabilité interne des secteurs.

Nous obtenons de la même manière la décomposition classique de l'inertie de l'ensemble des ménages en considérant le niveau d'agrégation, la Parcelle :

inertie de l'ensemble des ménages = inertie inter-parcelles + inertie intra-parcelles

3. 1. 3. DECOMPOSITION DE L'INERTIE SUIVANT UN SYSTEME D'ECHELLES

3. 1. 3. 1. Principe

A partir de ce principe de décomposition pour un seul niveau d'agrégation, nous démontrons (cf annexe 2.2) que l'expression de l'inertie du nuage de points N(C) sur lequel agissent deux partitions emboîtées A et B nous donne :

$$In(N(C)) = In.inter(A) + In.intm(BA) + In.intra(B)$$

où In.inter(A) est l'inertie inter-classes classique associée à la partition la moins fine A.

In.intm(BA) est "l'inertie intermédiaire" c'est-à-dire l'inertie du nuage des centres de gravité des unités b appartenant à une même unité a. Il s'agit d'une **inertie intra-classes dont les éléments de base, B, sont déjà des agrégats** et qui sont partitionnés suivant un niveau d'agrégation supérieur A. L'analyse du système d'échelles va reposer sur l'étude de cet élément intermédiaire.

In.intra(B) est l'inertie intra-classes classique associée à la partition B sur l'ensemble C.

et l'on a :

$$In.inter(B) = In.inter(A) + In.intm(BA)$$

$$In.intra(A) = In.intm(BA) + In.intra(B)$$

$$In.intm(BA) = In.inter(B) - In.inter(A) = In.intra(A) - In.intra(B)$$

où In.inter(B) est l'inertie inter-classes associée à la partition B sur l'ensemble C.
In.intra(A) est l'inertie intra-classes associée à la partition A sur l'ensemble C.

ce qui aboutit au schéma suivant :

$$In(N(C)) = \underbrace{In.inter(A)}_{In.inter(B)} + In.intm(BA) + \overbrace{In.intra(B)}^{In.intra(A)}$$

Remarque :

- L'inertie inter-classes associées à la partition A sur C peut être considérée comme l'inertie intra-classes associées à la partition réduite à une seule classe { C} sur A. De ce fait la **décomposition de l'inertie peut être envisagée comme la somme des inerties intra-classes entre deux niveaux d'agrégation** (inerties intermédiaires).

$$In(N(C)) = In.intm(AC) + In.intm(BA) + In.intm(CB)$$

3. 1. 3. 2. Généralisation

Soit P un ensemble de partitions emboîtées sur C,

$P = \{ P_0, P_1, ..., P_n, ..., P_N\}$ où $P_0 = \{ \{c\} \mid c \in C\}$ et $P_N = \{C\}$ alors :

$$In(N(C)) = \sum_{1}^{N} In.intm(P_{n-1}P_n)$$

$$In.inter(P_n) = \sum_{n+1}^{N} In.intm(P_{n-1}P_n) \text{ et } In.intra(P_n) = \sum_{1}^{n} In.intm(P_{n-1}P_n)$$

où $In.inter(P_n)$ et $In.intra(P_n)$ sont les inerties inter et intra-classes associées à la partition P_n sur l'ensemble C.

On a donc :

$$In.intm(P_{n-1}P_n) = In.inter(P_{n-1}) - In.inter(P_n)$$

Remarque :

- $In.intm(P_{n-1}P_n)$ peut alors être mis en parallèle avec l'indice de niveau de la classification ascendante hiérarchique.

3. 1. 3. 3. Exemple

Si l'on dispose d'un système d'échelles c'est-à-dire d'une série de partitions emboîtées (Parcelle et Secteur) sur un ensemble d'unités statistiques de base (les ménages) schématisé par la figure suivante :

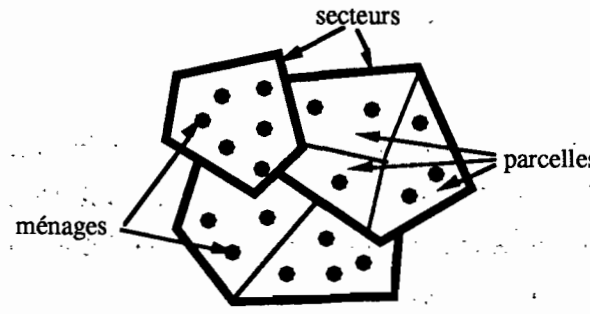


Figure n°18 : *Emboîtement Ménage-Parcelle-Secteur*

l'inertie totale de l'ensemble des ménages, In(M), se décompose de la manière suivante:

$$In(M) = In.Inter(S) + In.Intra(PS) + In.Intra(P)$$

où

- In.Inter(S) est l'inertie inter-secteurs classique qui mesure **la variabilité** c'est-à-dire l'écart **entre les secteurs**.
- In.Intra(PS) est l'inertie intra-secteurs sur l'ensemble des parcelles et mesure l'écart des parcelles (centres de gravité des ménages résidant sur une même parcelle) par rapport à leur secteur d'appartenance (centre de gravité des ménages ou des parcelles d'un même secteur). Elle exprime **la variabilité entre les parcelles à l'intérieur des secteurs** et traduit le passage entre deux niveaux d'agrégation.
- In.Intra(P) est l'inertie intra-parcelles classique sur l'ensemble des ménages et mesure **la variabilité des ménages à l'intérieur des parcelles** c'est-à-dire l'écart des ménages par rapport à leur parcelle.

Nous représentons cette décomposition de l'inertie de l'ensemble de ménages par :

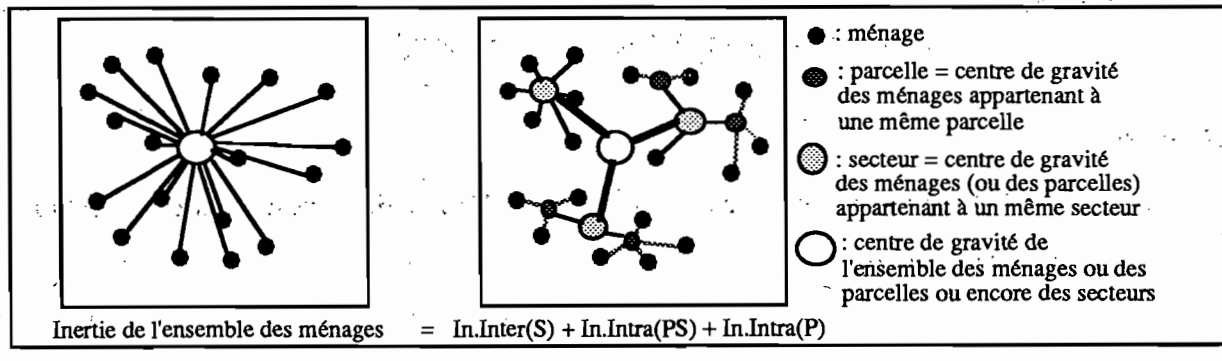


Figure n°19 : *Décomposition de l'inertie suivant l'emboîtement Ménage-Parcelle-Secteur*

Remarque :

- en sommant : In.Inter(S) + In.Intra(PS) = In.Inter(P)
  et In.Intra(PS) + In.Intra(P) = In.Intra(S)
  on obtient la décomposition classique de l'inertie de l'ensemble des ménages suivant un seul niveau d'agrégation, respectivement, la Parcelle et le Secteur.

Chapitre 3.2

ANALYSE SUR LES ELEMENTS DE LA DECOMPOSITION DE L'INERTIE

L'inertie pouvant être considérée comme une "**variance multivariée**", la méthode que nous proposons, revient à envisager une analyse de la variance décomposée suivant une série de partitions emboîtées, c'est-à-dire un système d'échelles, et adaptée aux techniques d'analyse multivariée pour des variables qualitatives. On rappelle que l'analyse de la variance[1], très utilisée en géographie, s'applique à l'explication des différenciations spatiales d'une variable de nature quantitative à partir d'une ou de plusieurs variables explicatives qualitatives définies par l'appartenance à une unité spatiale de niveau supérieur. Elle permet d'appréhender le rôle du niveau d'agrégation dans l'étude d'une variable quantitative.

Il s'agit pour cela de mettre en évidence les caractéristiques des diverses formes de variabilités obtenues dans la décomposition de l'inertie suivant un système d'échelles. Nous allons alors définir les tableaux associés aux éléments de la décomposition de l'inertie afin d'effectuer une analyse factorielle des correspondances appropriée au traitement de variables qualitatives. Nous présentons en annexe 2.3 une autre expression de la décomposition de l'inertie donnée d'après le terme général du tableau qui croise les variables entre elles c'est-à-dire du tableau assimilé à celui de Burt.

3. 2. 1. RAPPEL SUR L'ANALYSE FACTORIELLE DES CORRESPONDANCES

L'analyse factorielle des correspondances optimise la représentation de la dispersion des éléments afin de discriminer ceux qui se différencient le plus du profil moyen c'est-à-dire du centre de gravité du nuage étudié. La distance utilisée pour mesurer les ressemblances ou dissemblances entre les profils est la distance du chi-deux qui possède la propriété d'équivalence distributionnelle : si les classes sont homogènes c'est-à-dire s'il existe un regroupement d'éléments distributionnellement proches, il est sensiblement équivalent de réaliser l'analyse au niveau de ces unités et au niveau des unités de base. On peut ainsi, sous la condition d'homogénéité des classes, agréger les éléments d'une même classe,

[1][HAGGETT, 1968], [SANDERS,1989]

autrement dit effectuer un changement d'échelle, et conserver les structures de l'analyse du niveau le plus fin. "*l'intérêt de l'analyse des correspondances est que, du fait de l'équivalence distributionnelle, cette méthode est peu sensible au détail de la partition adoptée*"[1]. Une classe est dite homogène si "*les éléments qui la composent se ressemblent plus entre eux qu'ils ne ressemblent à l'ensemble des éléments*"[2].

Nous introduisons également la notion du modèle proposée par B.Escofier[3] et qui est une généralisation de l'analyse des correspondances. "*L'analyse des correspondances classique analyse l'écart entre un tableau de fréquence et un tableau modèle correspondant à l'hypothèse d'indépendance*". Ce modèle est le produit des marges du tableau de données." *L'analyse factorielle des correspondances se généralise à un modèle différent du modèle d'indépendance en supposant que les deux marges du tableau modèle sont égales à celle du tableau étudié*"[4]. La notion de modèle permet alors de généraliser la plupart des principes et des propriétés de l'analyse des correspondances ce qui va nous être utile par la suite. "*Pour appliquer un programme classique d'analyse des correspondances, il faut introduire le nouveau modèle et supprimer l'ancien*" c'est-à-dire le produit des centres de gravité des deux nuages duaux :

..." *données - modèle + produit des marges*" [4]

Nous précisons que le tableau des données et le tableau modèle doivent être de même dimension c'est-à-dire doivent être définis sur le produit des mêmes ensembles. De plus, ils doivent avoir les mêmes marges. Nous verrons alors que les tableaux associés aux éléments de cette décomposition ne peuvent être les tableaux de correspondances définis dans la deuxième partie. On considère uniquement les tableaux de centres de gravité et les tableaux assimilés à des tableaux de Burt associés. Les analyses factorielles appliquées à ces deux tableaux associés à un même niveau d'analyse sont équivalentes.

Il s'agit alors de définir le rôle que jouent les éléments de la décomposition et de comprendre ce que peut apporter cette décomposition pour l'analyse générale du nuage obtenue à partir de l'analyse factorielle des correspondances classique.

Nous allons réaliser une série d'analyses inter-classes et intra-classes pour lesquelles nous donnerons des éléments d'interprétation. Nous nous attachons particulièrement au **terme intermédiaire de la décomposition de l'inertie sur lequel repose le principe de l'analyse du système d'échelles.**

---

[1][BENZECRI, 1973b] T IIAn°1 p21.
[2][GRASLAND, 1991]
[3][ESCOFIER, 1984]
[4][ESCOFIER, 1988] p204.

3. 2. 2. L'ANALYSE INTER-CLASSES

3. 2. 2. 1. Le principe

L'analyse inter-classes étudie l'écart entre le profil moyen d'une classe a de A et le profil moyen du nuage de points C : un élément c d'une classe a est donc assimilé au centre de gravité de la classe et n'est étudié qu'au travers de celui-ci. Elle traduit la dispersion ou la variabilité entre les classes définissant le niveau d'agrégation et est réalisée sur le nuage des centres de gravités de chaque classe. Le modèle considéré est le modèle d'indépendance observé sur les deux ensembles A et J.

3.2.2. 2. Construction du tableau

On réalise une analyse des correspondances sur un des trois tableaux associé à un niveau d'agrégation (cf figure n°13) exposé dans le processus d'agrégation (§ 2.3.2.) : le tableau de correspondances croisant les classes constituant le niveau d'agrégation avec l'ensembe des variables ; le tableau croisant les variables entre elles (le tableau assimilé à un tableau de Burt) ; le tableau des centres de gravité (croisant les unités statistiques élémentaires analysées au travers du profil moyen de leur classe d'appartenance avec l'ensemble des variables) (cf annexe 2.5).

3. 2. 2. 3. L'interprétation

On met en évidence les associations entre les variables définissant ainsi des groupes d'unités statistiques ("élémentaires " ou "agrégats"). On en déduit les structures communes : si celles-ci sont très peu marquées, on obtient un mélange de tendances diverses, sinon ces structures peuvent être influencées de manière prépondérante par un groupe marginal particulièrement typé.

3. 2. 3. L'ANALYSE INTRA-CLASSES

3. 2. 3. 1. Le principe

Parce que l'analyse inter-classes est usuelle, la plupart des études portant sur la décomposition de l'inertie privilégient l'analyse intra-classes. Elle est réalisée à partir des tableaux ternaires ou des tableaux disjonctifs complets qui, de par leur construction, sont partitionnés (cf annexe 2.4).

Son intérêt est d'**étudier la dispersion des éléments**, constituant le nuage, **à l'intérieur de leur classe d'appartenance**. De ce fait, les profils des éléments sont analysés non pas par rapport au profil moyen du nuage mais par rapport au profil moyen de la classe. Le modèle de référence n'est plus le modèle d'indépendance observé sur les deux ensembles C et J mais celui observé sur C_a et J, où C_a est l'ensemble des unités statistiques c appartenant à une même classe a. Il traduit exactement la dispersion entre les classes (on vérifie que les marges de ce modèle sont les mêmes que celles du tableau initial) et n'est rien d'autre que le tableau associé à l'analyse inter-classes. L'inertie inter-classes est par conséquent supprimée afin de ne considérer que les variations internes aux classes.

3. 2. 3. 2. Construction du tableau

Ainsi si l'on considère le niveau d'agrégation, B, le tableau qui traduit l'inertie intra-classes a pour terme général (cf annexe 2.6) :

$$HC_{cj} - \frac{HB_{bj}}{n_b} + (HC_c * HC_j)$$

où HC(CJ) est le tableau de données dont la structure de base, disjonctive complète, ne peut être décomposée.

$\frac{HB_{BJ}}{n_b} = HC°_{cj}$ est le tableau modèle qui correspond au tableau des centres de gravités des classes définissant le niveau d'agrégation et n_b est le nombre d'unités c (les ménages) de la classe b (la parcelle).

On a de la même manière pour l'inertie inter-classes associées au niveau d'agrégation A :

$$HC_{cj} - \frac{HA_{aj}}{n_a} + (HC_c * HC_j)$$

Remarque :

- Lorsque le tableau de données HC_{CJ} est un tableau disjonctif complet, l'analyse intra-classes est une analyse des correspondances multiples conditionnelle[1].
- Les tableaux HC_{CJ}, $HC°_{CJ} = \frac{HB_{BJ}}{n_b}$, sont définis sur le même produit des ensembles C*J. On construit également les tableaux qui leur sont associés et qui croisent les variables qualitatives

[1] [ESCOFIER,1987b]

deux à deux c'est-à-dire les tableaux assimilés à des tableaux de Burt et définis sur le produit J*J (cf annexe 2.6).

3. 2. 3. 3. L'interprétation

Effectuer une analyse factorielle des correspondances sur un tel tableau revient à effectuer l'analyse simultanée des sous-nuages des profils-lignes (sur lequel a été opérée la partition) recentrés à leur origine. En effet, "*le nuage associé à l'inertie intra-classes est construit en disposant dans* R^J *autour du centre de gravité du nuage total les vecteurs différences entre chaque élément de* I_t *et le centre de sa classe*"[1].

Il s'agit d'étudier les écarts entre les profils d'une même classe et "*en effectuant une analyse factorielle des correspondances des sous-tableaux* (définis par les sous-nuages) *de mesurer quelle part chaque classe est capable de restituer de la structure définie en commun*"[2]. On élimine l'effet dû à la partition c'est-à-dire à l'agrégation.

3. 2. 4. ANALYSE DE "L'INERTIE INTERMEDIAIRE"

3. 2. 4. 1. Le principe

Nous avons vu que "l'inertie intermédiaire" exprime une inertie intra-classes dont les éléments de base constituant le nuage sont déjà des agrégats plus ou moins homogènes.

In.Intrm(BA) est **l'inertie intra-classes associées au niveau d'agrégation supérieur A sur l'ensemble des classes b de B** c'est-à-dire sur l'ensemble des centres de gravité des unités de base c appartenant à une même classe b.

L'analyse associée à cette inertie rejoint les principes de l'analyse intra-classes et doit traduire la variabilité des classes constituant le niveau d'agrégation inférieur à l'intérieur de leurs classes d'appartenance qui définissent le niveau d'agrégation supérieur.

---

[1] [BENZECRI, 1983] p357.
[2] [CAZES *et al*, 1988] p40.

3. 2. 4. 2. Construction du tableau

On construit (cf annexe 2.7) le tableau qui exprime cette variabilité de la manière suivante : au tableau de données traduisant le niveau d'agrégation inférieur, on retranche le tableau "modèle" correspondant au niveau d'agrégation supérieur ; celui-ci traduit l'hypothèse d'indépendance pour chaque sous-nuage constitué par les classes. Puis on ajoute le produit des marges $(HB_b * HB_j)$ qui correspond à l'hypothèse d'indépendance classique afin de pouvoir effectuer le programme usuel de l'analyse des correspondances. (On vérifie que tous ces tableaux ont les mêmes marges).

Si l'on considère les deux niveaux d'agrégation B et A où A englobe B, le terme général est :

$$\frac{HB_{bj}}{n_b} - \frac{HA_{aj}}{n_a} + (HB_b * HB_j)$$

où n_b est le nombre d'unités c (les ménages) de la classe b (la parcelle).

Le tableau de données $\frac{HB_{BJ}}{n_b} = HC°_{CJ}$, est le tableau des centres de gravités des classes définissant le niveau inférieur B.

n_a est le nombre d'unités c (les ménages) de la classe a (le secteur). Le tableau modèle $\frac{HA_{AJ}}{n_a} = HC°°_{CJ}$ est le tableau des centres de gravités des classes définissant le niveau supérieur A.

et l'on vérifie que les marges des tableaux sont identiques :

$$HC_c = HB_b = HA_a \qquad \text{et} \qquad HC_j = HB_j = HA_j$$

Par conséquent, l'analyse de ce tableau réalise au préalable une analyse inter-classes du nuage $N(C_a)$ partitionné par B_a. C'est finalement l'analyse intra d'une analyse inter.

Remarques :

- Les tableaux $HC°_{CJ} = \frac{HB_{BJ}}{n_b}$ et $HC°°_{CJ} = \frac{HA_{AJ}}{n_a}$, sont définis sur le produit des mêmes ensembles C*J. On construit les tableaux assimilés à des tableaux de Burt et définis sur le produit J*J qui leur sont associés et l'on démontre que les analyses réalisées sur des deux types de tableaux sont équivalentes (cf annexe 2.7).

- L'analyse des correspondances associée à "l'inertie intermédiaire" peut être alors considérée comme une généralisation de l'analyse des correspondances multiples conditionnelle c'est-à-dire une analyse intra-classes appliquée non pas à un tableau disjonctif complet mais à une juxtaposition de tableaux de contingence.

- A la différence de l'analyse intra-classes effectuée sur un tableau ternaire, les sous-tableaux définis par les sous-nuages, considérés dans la décomposition de l'inertie pour une série de partitions emboîtées, sont de dimensions différentes.

3. 2. 4. 3. L'interprétation

L'examen des formules barycentriques présentées en annexe 2.7 nous permettent d'interpréter l'analyse des correspondance effectuée sur le tableau HB*(B,J) : celle-ci dégage les structures qui sont communes aux classes de la partition définie par le niveau d'agrégation le plus fin, et internes aux classes associées à la partition la moins fine correspondant au niveau d'agrégation supérieur. On peut ainsi saisir sans ambiguïté **les structures induites pour un niveau d'agrégation donné et qui vont être perdues dans le passage vers un niveau supérieur**. Il s'agit alors uniquement de certaines structures spécifiques du niveau inférieur données par l'analyse inter-classes qui vont disparaître lors du passage vers une échelle plus globale, et non pas de toutes les structures dégagées sur ce même niveau.

3. 2. 5. CONSEQUENCES DE LA STRUCTURE HIERARCHIQUE SUR LES DEUX NUAGES

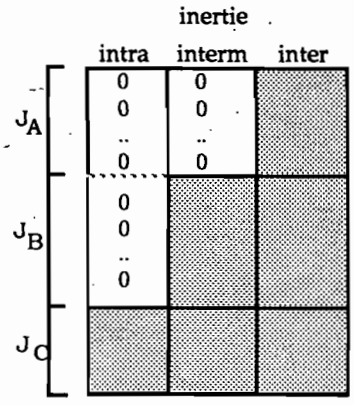
3. 2. 5. 1. Sur le nuage de individus

Une des conséquences principales de la structure hiérarchique sur le nuage des individus est le cas de classes constituées d'un seul élément. Il est évident que l'inertie intra-classe de cette classe est nulle. Le problème réside alors dans la différenciation entre les classes non réduites à un seul élément et parfaitement homogènes (tous les éléments d'une classe sont semblables) et les classes composées d'un seul élément.

3. 2. 5. 2. Sur le nuage des modalités

Du fait de la répétition des informations des niveaux supérieurs à un niveau inférieur, tous les individus d'une classe possèdent nécessairement les mêmes modalités spécifiques à la population engendrant la partition. Pour trois niveaux d'observation emboîtés A, B et C auxquels correspondent

respectivement les groupes de variables J_A, J_B, J_C, l'inertie intra-classes associées au niveau de définition des modalités est nulle[1] ce qui est représenté par :

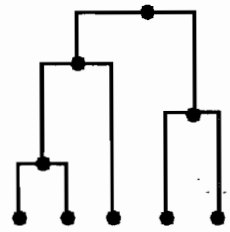


De ce fait, ce sont essentiellement les modalités spécifiques du niveau d'agrégation puis à un degré moindre celles du niveau immédiatement inférieur (etc...) qui contribuent à l'inertie inter-classes.

3. 2. 6. UN PROCESSUS DE CLASSIFICATION

Dans la mesure où l'étude d'un système d'échelles par l'analyse des correspondances repose sur la décomposition de l'inertie pour une série de partition emboîtées, on peut, sans formalisation mathématique, rapprocher le processus d'un système d'échelles de celui d'une classification ascendante hiérarchique mise à part que :

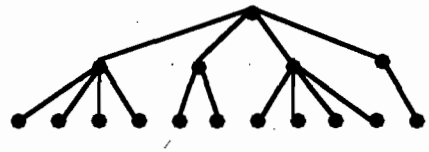
- **La classification ascendante hiérarchique est un processus automatique qui cherche à organiser des données en créant un lien** (défini suivant un critère d'agrégation de la maximisation de l'inertie inter-classes). Il s'agit d'un arbre connexe binaire :



- **Un système d'échelles est un processus "naturel" qui explique une organisation des données déterminée par l'existence d'un lien hiérarchique** tel que nous l'avons défini (c'est-à-dire

[1][PIRON, 1990]

traduit par une série de partitions emboîtées) et auquel on cherche à donner un sens en terme d'homogénéité. Il s'agit d'un arbre connexe fourchu :



Le rapprochement s'effectue non pas dans la formation des classes mais dans leur interprétation. Pour cela, deux aides sont données par :

- La série d'analyses inter-classes qui joue le rôle de l'écart d'une classe au centre du nuage. On recherche les associations de variables qui sont, non pas responsables de la formation des classes, mais qui donnent les éléments structurant des classes du niveau considéré.

- La série d'analyses intra-classes qui joue le rôle de l'expression des dipôles relatifs aux nœuds supérieurs. Elle met en évidence, non pas ce qui oppose l'ainé et le benjamin d'une classe, mais les éléments qui diffèrent et se regroupent dans les mêmes classes pour un niveau considéré.

De plus, nous avons vu que la décomposition de l'inertie pouvait être envisagée comme la somme des inerties intra-classes entre deux niveaux consécutifs et ce que nous avons appelé inertie "intermédiaire". Chaque inertie intermédiaire dégagée joue le rôle de l'indice de niveau, mise à part qu'elle ne décroît pas au fur et à mesure que les partitions sont plus fines puisque :

- "**L'inertie intermédiaire**" est un critère qui **détermine la perte d'inertie** obtenue en passant d'un niveau à un autre.

- **L'indice de niveau** est un critère qui **minimise la perte d'inertie** réalisée en passant d'une partition à une autre c'est-à-dire d'un niveau à un autre.

L'examen des "inerties intermédiaires" doit permettre de décider du niveau à la fois le moins fin et qui rend compte au mieux d'une homogénéité globale. Leur histogramme est alors mis en parallèle avec celui des indices de niveau donné dans la classification ascendante hiérarchique.

Pour l'interprétation de la classification ascendante hiérarchique, le principe est de s'intéresser à un ensemble de classes pour une partition fixée et à la compréhension de leur organisation alors que pour le système d'échelles, le principe est de saisir un ensemble de partitions c'est-à-dire de niveaux et d'étudier leur influence pour un phénomène donné.

AIDES A L'INTERPRETATION DU SYSTEME D'ECHELLES

L'étude du système d'échelles repose sur le fait que nous disposons de plusieurs niveaux d'observation et donc de plusieurs niveaux d'analyse. Nous avons cherché d'une part, à les caractériser et d'autre part, à mettre en évidence les relations d'un niveau à un autre en approfondissant la décomposition de l'inertie dans le cas d'une série de partitions emboîtées.

Ainsi, on aboutit à deux séries d'analyses, inter et intra-classes, qui expliquent la formation des niveaux et leurs disparités internes. Pour une vision globale des phénomènes schématisée par la figure n° 20, cela signifie que l'on dégage **les structures communes des classes correspondant au niveau le plus fin considéré, d'agrégation ou non,** et qui se modifient progressivement au fur et à mesure que l'on passe à des niveaux supérieurs d'agrégation.

Par la suite, on cherche à définir un niveau pertinent qui traduit les phénomènes à une échelle globale. Autrement dit, il s'agit de décider d'une partition qui soit la plus homogène afin que son analyse ne soit pas perturbée par rapport à celle du niveau initial. Puis nous nous intéressons à la détermination du niveau d'intervention des différentes variables.

3. 3. 1. L'ANALYSE DES NIVEAUX

L'analyse des niveaux revient à réaliser une succession d'analyses inter-classes afin de dégager les structures spécifiques de chaque niveau et de saisir ainsi l'influence des niveaux sur la perception des phénomènes. On effectue une analyse des correspondances sur un des tableaux associés à un niveau d'analyse (fig. n°13) qui sont exposés dans le processus d'agrégation (§2.3.2.) :

- **au niveau le plus fin,** C, (le Ménage par exemple), on effectue une analyse sur le tableau disjonctif complet HC(C,J) c'est-à-dire sur le tableau de données non agrégées (fig. n°20a) (ou encore sur le tableau de Burt BC(J,J)). A ce niveau d'analyse, l'information n'est pas transformée et l'analyse nous fournit une organisation du maximum d'information. On obtient des profils des unités statistiques de base (les ménages).

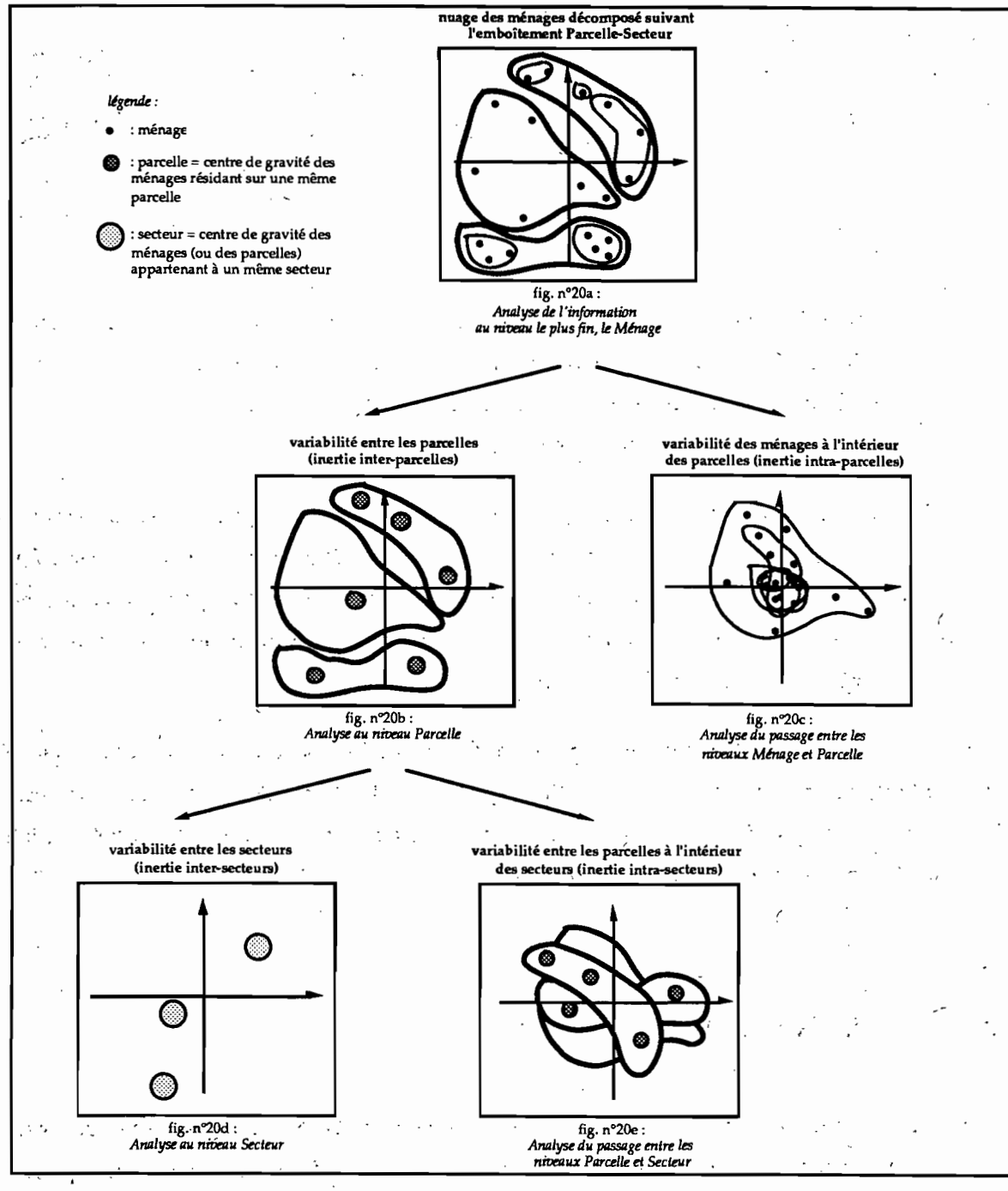


Figure n°20 : *Analyse de la décomposition de l'inertie sur le nuage des ménages suivant l'emboîtement Ménage-Parcelle-Secteur*

- **au niveau intermédiaire**, B, (la Parcelle), une unité statistique de base (un ménage) est analysée au travers de la classe (la parcelle), constituant ce niveau intermédiaire, à laquelle elle appartient. On réalise une analyse des correspondances sur le tableau, HB(B,J), croisant l'ensemble des classes avec celui des variables (ou encore sur les tableaux BB(J,J) ou $HC^{\circ}(C,J)$) (fig. n°20b). On obtient des profils de classes (de parcelles) définissant ce premier niveau d'agrégation.

- **au niveau supérieur,** A, (le Secteur), une unité statistique plus fine (un ménage mais aussi une parcelle) est analysée au travers de la classe (le secteur), constituant le niveau supérieur, à laquelle elle appartient. La classe (le secteur) est représentée par le centre de gravité des unités statistiques de base (les ménages) qui est identique à celui des unités statistiques déjà agrégées (les parcelles). On effectue une analyse des correspondances sur le tableau, HA(A,J), croisant l'ensemble des classes avec celui des variables (ou encore sur les tableaux BA(J,J),HC°°(C,J) ou HB°(B,J)) (fig. n°20d). On obtient des profils de classes (de secteurs) constituant le niveau supérieur d'agrégation.

L'analyse du niveau d'agrégation est fréquemment et implicitement réalisé puisqu'il s'agit de l'analyse du tableau issu d'une bande de tableau de Burt comme par exemple une matrice d'information spatiale qui croise une variable géographique (la région par exemple) avec un ensemble de variables. Elle met en évidence les associations ou les différenciations entre les classes définissant le niveau d'agrégation c'est-à-dire la variabilité entre ces classes.

3. 3. 2. L'ANALYSE DU PASSAGE D'UN NIVEAU A UN AUTRE

L'analyse des niveaux d'agrégation se conçoit comme une succession d'analyses inter-classes qui examinent les structures communes aux classes. Mais puisque nous nous intéressons non seulement aux structures propres d'un niveau mais également aux structures perdues dans le changement de niveau, pour ce dernier aspect, on admet qu'il s'agit également d'une succession d'analyses intra-classes qui définissent les caractéristiques des passages entre les niveaux. On cherche **à éliminer l'effet du à l'agrégation**. On représente une unité statistique ("élémentaire" si elle est prise au niveau d'analyse de base, (le Ménage), ou une classe si elle est prise à un niveau déjà agrégé), non pas par son écart à la moyenne générale de la population d'étude, comme dans une analyse classique, mais par son écart à la moyenne de la population appartenant à la même unité spatiale que lui.

Par conséquent, l'analyse du passage entre deux niveaux revient à étudier la dispersion des unités statistiques à l'intérieur de leur classe d'appartenance. On a deux cas de figures :

- **les unités statistiques sont définies par le niveau d'analyse de base** C (le Ménage), et nous avons une structure de base (disjonctive complète) ne pouvant être décomposée. Pour étudier le passage entre les niveaux C (Ménage) et B (Parcelle), on réalise une analyse des correspondances sur le tableau suivant :

$$HC_{cj} - \frac{HA_{aj}}{n_a} + (HC_c * HC_j)$$

Cette analyse (fig.n°20c) recentre les unités c (les ménages) par rapport au centre de gravité de leur classe (la parcelle) à laquelle ils appartiennent. Elle fournit ainsi **les indicateurs de disparité des ménages à l'intérieur des parcelles.**

- **les unités statistiques correspondent à des classes** qui définissent le niveau d'agrégation le plus fin considéré B (la Parcelle): pour étudier le passage entre les niveaux d'agrégation B et A (Parcelle et Secteur), on effectue une analyse des correspondances sur le tableau :

$$\frac{HB_{bj}}{n_b} \cdot \frac{HA_{aj}}{n_a} + (HB_b * HB_j)$$

Cette analyse (fig.n°20e) recentre les unités statistiques déjà agrégées b (les parcelles) par rapport à la classe (le secteur) auquel elles appartiennent. Par exemple, les parcelles sont représentées par le centre de gravité des ménages y résidant et les secteurs par le centre de gravité des parcelles ou des ménages lui appartenant. On obtient ainsi **les indicateurs de disparité entre les parcelles à l'intérieur des secteurs.**

On envisage ainsi les passages successifs d'un niveau à un autre. L'analyse du passage entre deux niveaux d'agrégation fournit les caractéristiques de différenciation au sein d'une même classe définie par le niveau supérieur. Elle permet de **saisir les structures induites par le niveau inférieur et qui sont perdues dans le changement d'échelle.**

3. 3. 3. CRITERE D'HOMOGENEITE DES NIVEAUX

La succession d'analyses des niveaux d'agrégation et des passages entre ces niveaux aboutit à l'analyse du système d'échelles. La figure n°20 schématise le processus d'analyse de la décomposition de l'inertie pour une série de partitions emboîtées. Dans le cadre d'analyses factorielles, nous dirons qu'**un niveau** (le Secteur par exemple) **est globalement homogène (hétérogène) si les classes** (les secteurs) **qui le composent sont faiblement (fortement) différenciés** compte tenu des phénomènes mis en évidence par le plan factoriel. Cela signifie que les unités statistiques (les parcelles ou les ménages) qui appartiennent à une même classe (un secteur) ont des profils semblables. On peut ainsi représenter les différents cas d'homogénéité des niveaux :

- Si **l'analyse du niveau supérieur rend compte au mieux des mêmes structures** (mêmes axes factoriels) que celle du niveau inférieur, alors l'analyse du passage entre ces deux niveaux dégage une perte minimale d'information due à l'agrégation mais qui ne bouleverse pas l'organisation des données au niveau le plus fin : l'inertie intra-classes est faible et **les classes du niveau supérieur sont globalement homogènes** (cf figure n°21a).

  Dans ce cas, on peut ramener **l'étude au niveau supérieur d'agrégation** (le Secteur) **qui conserve alors la stabilité du phénomène observé au niveau inférieur** (la Parcelle).

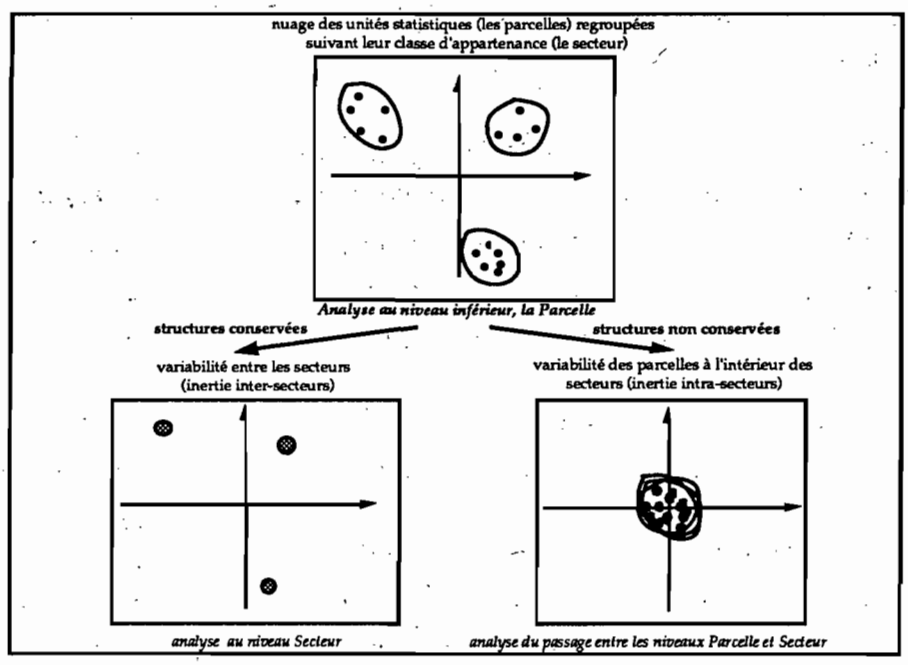


Figure n°21a : *Niveau d'agrégation globalement homogène (classes homogènes)*

- Si **l'analyse du niveau supérieur, en revanche, ne se structure pas comme celle du niveau inférieur,** alors l'analyse du passage rend compte d'une même organisation du nuage de points que celle du niveau inférieur : l'inertie intra-classes est élevée et **le niveau supérieur est globalement hétérogène** (cf figure n°21b).

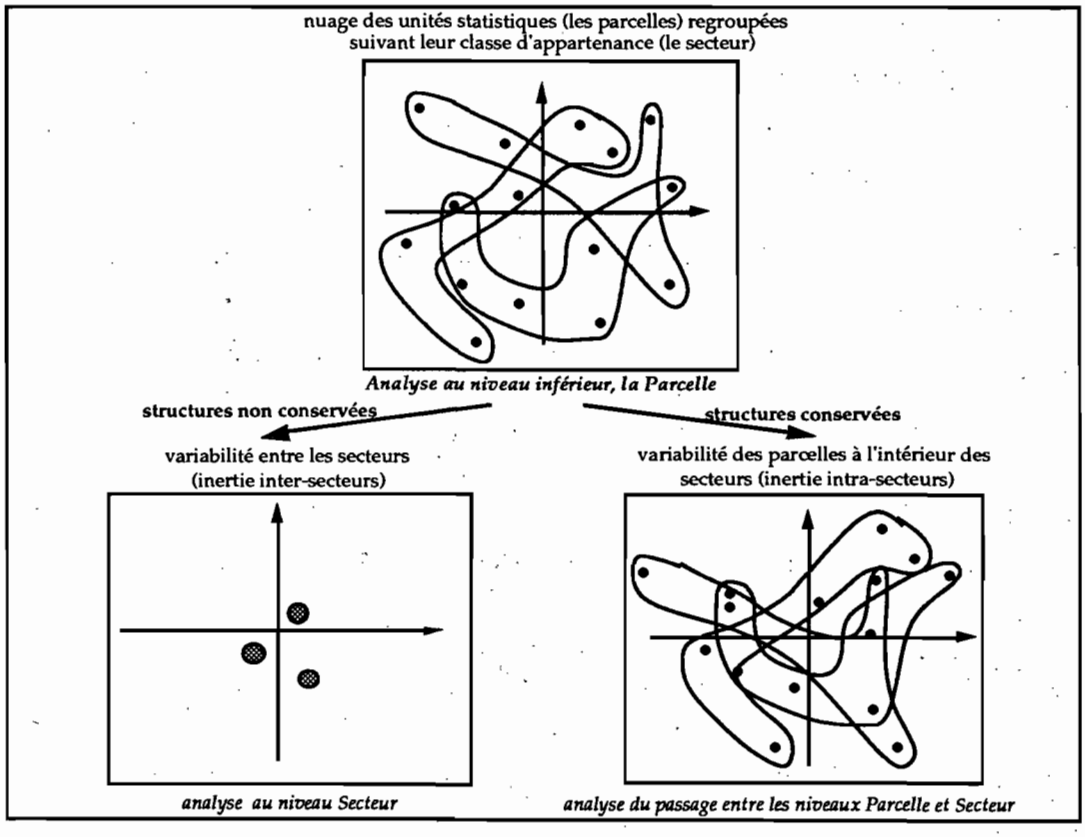


Figure n°21b : *Niveau d'agrégation globalement hétérogène (classes hétérogènes)*

Dans ce cas, les analyses aux deux niveaux distinctes, la Parcelle et le Secteur, apportent deux points de vue différents.

Par conséquent, si l'analyse inter-classes est implicitement utilisée dès que l'on traite de données agrégées, l'analyse intra-classes met en évidence ce qui s'est passé dans le changement d'échelle en dégageant la nature de la transformation de l'information, que ce soit un phénomène minime (faible variabilité) mais intéressant ou un phénomène important (forte variabilité). L'analyse correspondant à l'inertie la plus faible complète la connaissance du phénomène étudié.

On peut appréhender l'homogénéité des niveaux du système par l'examen de ces deux types d'analyses. Est-il alors possible de donner **une mesure du critère d'homogénéité des niveaux ?**

A.S.FOTHERINGHAM et D.W.S.WONG[1] posent le problème de résultats variables d'une part suivant les échelles, d'autre part suivant les découpages pour une échelle donnée. On est tenté de vouloir donner une mesure du critère d'homogénéité des niveaux par exemple à partir des taux d'inertie: $\frac{\text{inertie inter}}{\text{inertie totale}}$ ou $\frac{\text{inertie intra}}{\text{inertie totale}}$. Cependant, il est délicat de vouloir obtenir une mesure du critère d'homogénéité pour un niveau donné à partir de l'inertie.

En effet, dans le cadre d'analyses multivariées, ces taux dépendent du découpage (nombre de classes associées au niveau) et du nombre de variables de l'étude : en analyse des correspondances multiples *"la part d'inertie due à une variable est fonction croissante du nombre de modalités de la variable"*[2]. Cela signifie que l'inertie inter-classes associées à un niveau d'agrégation (qui est égale à l'inertie associée à un tableau de contingence, ou encore à une bande d'un tableau de Burt), dépend du nombre de variables de l'étude et du nombre de modalités de la variable de partition, c'est-à-dire du nombre de classes du niveau d'agrégation.

Dans le cas d'un système d'échelles, c'est-à-dire d'un ensemble de niveaux emboîtés, il est évident que plus le nombre de classes de ce niveau est élevé, plus ces classes sont homogènes (faible inertie intra-classes). Soit encore plus on agrège, plus les niveaux sont globalement hétérogènes.

$$In(N(C)) > In(N(B)) > In(N(A))$$

On rappelle alors que :

$$In(N(C)) = In.inter(A) + In.intm(BA) + In.intra(B)$$

et d'une manière plus générale $In(N(C)) = \sum_{1}^{N} In.intm(P_{n-1}P_n)$

[1][FOTHERINGHAM *et al*, 1991]
[2][LEBART *et al*, 1977] p134.

Si les inerties inter-classes augmentent au fur et à mesure que les partitions sont plus fines, les inerties intermédiaires ne peuvent répondre à un ordre préétabli puisque la série de partitions emboîtées du système d'échelles est imposée et ne vérifie pas *a priori* un critère d'homogénéité. Par conséquent, si l'inertie intermédiaire est faible, (peu de variation dans le passage d'un niveau à un autre), on peut ramener l'étude au niveau supérieur d'agrégation qui conserve alors la stabilité du phénomène observé au niveau inférieur. Les classes sont globalement homogènes par rapport aux classes du niveau inférieur.

Dans le cas de découpages obtenus à une même échelle pour un même nuage de points, il existe des variabilités différentes. C'est ce qui peut permettre de réfléchir à la pertinence des découpages. Les taux d'inertie diffèrent et les analyses des niveaux et du passage entre les niveaux sont différentes. Cela se conçoit facilement si nous considérons deux systèmes d'échelles de découpage identique à celui de l'étude:

- le premier est obtenu à partir de partitions issues d'une classification ascendante hiérarchique, A', B' sur C.
  On a : $In(N(C)) > In(N(B')) > In(N(A'))$

- le deuxième est obtenu à partir de partitions aléatoires, A", B" sur C.

  On a : $In(N(C)) > In(N(B")) > In(N(A"))$

et l'on peut très bien avoir : $In(N(A')) > In(N(B"))$

Du fait que le critère d'agrégation de la classification ascendante hiérarchique est fondé sur la minimisation de l'inertie intra-classes, on peut uniquement comparer les taux entre les systèmes d'échelles pour un même nuage de points et dire que le système d'échelles obtenu à partir de la classification ascendante hiérarchique rend compte au mieux de niveaux homogènes en fonction du critère d'agrégation choisi et l'on alors :

$$In(N(A')) > In(N(A))$$

$$In(N(B')) > In(N(B))$$

Il faudrait maintenant envisager de faire des tests statistiques pour connaitre la part due au découpage et ainsi essayer d'en déduire une mesure d'homogénéité.

3.3. 4. NIVEAU D'INTERVENTION DES VARIABLES

Après s'être intéressés aux différentes perceptions d'un phénomène donnés par l'étude des axes factoriels suivant les niveaux d'analyse, nous cherchons, ici, à déterminer le mode et l'intensité d'intervention de chaque variable, niveau par niveau. Il s'agit donc d'une étude ponctuelle de la variable considérée dans le système d'échelles, indépendamment de toutes combinaisons de variables.

L'inertie d'un nuage de points N(C) est la somme des inerties des variables q qui, elles-mêmes, sont la somme des inerties de leurs modalités j, on a :

$$\mathbf{In(N(C)) = In(N(J)) = \sum_{q} In(q))}$$

où

$$In(q) = \sum_{j \in q} In(j)$$

Compte tenu de la décomposition de l'inertie, on peut décomposer l'inertie d'une variable q et d'une modalité j suivant le même principe :

$$In(N(q)) = In.inter(A,q) + In.intm(BA,q) + In.intra(B,q)$$

et

$$In(N(j)) = In.inter(A,j) + In.intm(BA,j) + In.intra(B,j)$$

Ceci autorise alors à effectuer des pourcentages en ligne ou en colonne de l'inertie décomposée d'une modalité j ou d'une variable q.

Pour définir l'impact d'une variable q (ou d'une modalité j) dans une étape du changement d'échelle, nous considérons deux indices de contribution :

- le pourcentage en colonne pour chaque étape du changement d'échelle:

  - étude du passage, entre B et A par exemple :

$$CTR_1(BA,q) = \frac{In.intm(BA,q)}{In.intm(BA)}$$

  - étude du niveau, B par exemple :

$$CTR_1(B,q) = \frac{In.inter(B,q)}{In.inter(B)}$$

  On obtient la contribution relative de la variable q à l'inertie considérée (inter, intermédiaire ou intra).

- le pourcentage en ligne pour chaque variable q associé à une étape :

  - étude du passage : $CTR_2(BA,q) = \frac{In.intm(BA,q)}{In(q)}$

  - étude du niveau : $CTR_2(B,q) = \frac{In.inter(B,q)}{In(q)}$

  On obtient la contribution du niveau de la variable q à l'inertie de la variable.

On calcule les mêmes indices pour une modalité j.

Si $CTR_1(BA,q)$ ou $CTR_2(BA,q)$ sont élevées, la variable q est un élément de **dispersion interne** des individus de B au sein des classes définies par le niveau supérieur A.

Si $CTR_1(B,q)$ ou $CTR_2(B,q)$ sont élevées, la variable q est un élément de **concentration** des individus au sein des classes définies par le niveau B.

Le premier indice, CTR_1, permet de définir le rôle que joue chaque variable dans le découpage du système d'échelles.

Le deuxième indice, CTR_2 , permet de définir le niveau sur lequel la variable intervient davantage. Il met en évidence le niveau d'intervention de la variable : plus l'inertie inter-classes associée à une partition est importante, plus la variable contribue à caractériser le niveau correspondant à la partition. Inversement, plus l'inertie intra-classes est élevée, plus la variable contribue à caractériser le niveau inférieur considéré dans le passage.

Puisque $CTR_2(CB,q) + CTR_2(BA,q) + CTR_2(A,q) = 100\%$

cela nous autorise à construire un graphique qui permet de visualiser directement le niveau pertinent pour l'étude d'une variable.

PROPOSITION D'UNE METHODE D'ANALYSE D'UN SYSTEME D'ECHELLES

La méthode d'analyse du système d'échelles que nous proposons permet d'identifier la nature de l'information et ses transformations éventuelles du au changement d'échelle. Nous sommes ainsi en mesure de pouvoir décider du niveau pertinent d'analyse compte tenu du phénomène étudié.

A partir de l'analyse du niveau le plus fin qui donne l'information maximale, il s'agit de dégager les structures spécifiques de chaque étape du système d'échelles, schématisée par la figure suivante :

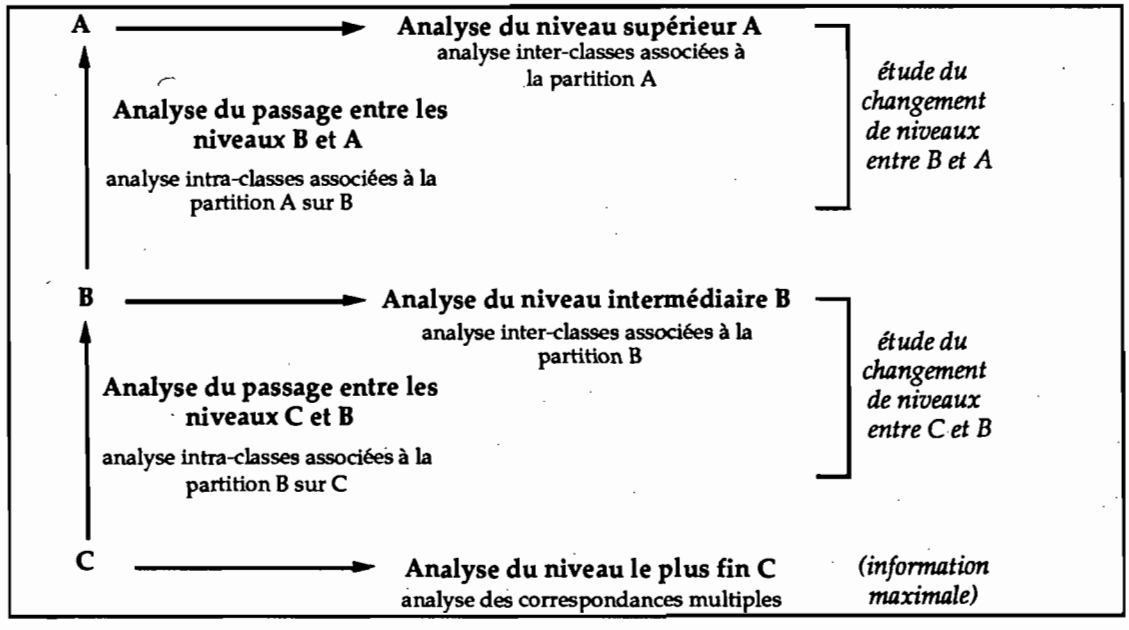


Figure n°22 : *Analyse d'un système d'échelles*

L'étude du changement d'échelle nécessite alors une double analyse:

- l'analyse du niveau pour identifier ce qui est conservé ou transformé de l'information initiale à un niveau supérieur;
- l'analyse du passage entre deux niveaux pour identifier la nature de l'information perdue dans le changement d'échelle.

Les deux types d'analyses sont indispensables pour le traitement de données agrégées. Cependant si elles permettent d'apprécier un niveau d'homogénéité globale, elles ne peuvent donner un indice d'homogénéité.

Quatrième Partie

UNE APPLICATION DE LA METHODE D'ANALYSE DU SYSTEME D'ECHELLES

L'analyse du système d'échelles doit permettre d'analyser les formes d'organisation spatiale et sociale qui caractérisent les quartiers périphériques de Ouagadougou.

Les principaux objectifs de l'étude sur l'espace périphérique de Ouagadougou sont "*d'analyser le processus de production de la ville, des formes d'organisation spatiale et sociale qui la caractérisent à partir d'une analyse des pratiques et des stratégies des différents agents de cette production : la population citadine, les pouvoirs publics et locaux et les intervenants extérieurs. Il s'agit de mettre en rapport les desseins politiques sur la ville et les pratiques sociales et résidentielles des citadins*"[1]. Au sein de cette problématique, on se demande d'une manière générale :

- Quelles sont les principaux paramètres sociologiques et fonciers de l'organisation spatiale de la périphérie de Ouagadougou ?

- Existe-t-il des regroupements sociologiques au sein de l'espace urbain ? sont-ils liés à la politique urbaine ?

Une des principales dimensions de cette politique urbaine est le lotissement systématique des quartiers spontanés de la périphérie. Ces espaces déjà ou prochainement restructurés sont le "*lieu privilégié pour la rencontre et l'affrontement des diverses logiques d'appropriation et de contrôle de l'espace urbain, lequel s'y révèle avec une netteté accusée comme un enjeu de rapports de pouvoir*".[1]

- Y a-t-il et quels sont les effets des opérations d'aménagement sur le comportement des ménages ?

Par conséquent, **il s'agit de mettre l'accent sur les processus mis en œuvre par les citadins dont les pratiques et les stratégies peuvent produire un espace socialement déterminé.**

[1][LE BRIS *et al*, 1990]

CONTEXTE EXPERIMENTAL : ENQUETE SUR OUAGADOUGOU

L'étude de l'organisation spatiale de la périphérie de Ouagadougou passe par d'une part la politique urbaine de l'état au travers de l'aménagement du tissu urbain et par une description des comportements de ménages au travers de leur accès au sol et au logement, de leur structure socio-économique et de leur intégration à la ville.

4. 1. 1. COMPOSITION DU SYSTEME D'ECHELLES

Il s'agit de déterminer les éléments qui interviennent dans le système d'échelles, c'est-à-dire les variables, la population statistique et les niveaux d'analyse (observés ou non).

Compte tenu des objectifs, nous nous intéressons aux deux niveaux d'observation du dispositif d'enquête :

- la **Parcelle** qui est le maillon de base d'une problématique foncière. Son étude doit permettre de définir les modalités d'accès au sol et au logement et de saisir les pratiques résidentielles ainsi que l'évolution des opérations de lotissement.

- le **Ménage** qui est l'unité "d'entrée" dans la réalité sociologique de la ville. L'intégration urbaine est perçue à travers l'arrivée à Ouagadougou et la mobilité intra-urbaine, informations issues de l'enquête rétrospective.

A ces deux niveaux correspondent deux groupes de variables :

- celles descriptives de la parcelle sur laquelle réside le ménage, soit 7 variables totalisant 24 modalités :

le nombre de ménages sur la parcelle	MEN	la superficie de la parcelle	SPF
le nombre de bâtiments	BAT	le mode d'acquisition	MAC
la date d'acquisition	DAC	les opérations de lotissements	OPL
le cout de construction	CTB		

- celles descriptives de tous les ménages résidant sur la parcelle, soit 14 variables totalisant 50 modalités :

le nombre de personnes dans le ménage	PRS	le sexe	SEX
l'âge	AGE	le statut matrimonial	STM
l'éthnie	ETH	le niveau d'instruction	SCL
l'activité	ACT	le statut d'occupation	STO
l'origine du chef de ménage	ORG	la date d'installation sur la parcelle	DRP
la date d'arrivée à Ouagadougou	DRO	le lieu d'arrivée à Ouaga	LRO
le nombre de migrations intra-urbaines	MIG	le type de ménage	SMN

Nous considérons alors la population statistique correspondant au niveau le plus fin compte tenu des variables sélectionnées, **la population des ménages**. Parmi les thèmes envisagés, l'intégration à la ville et l'aménagement du tissu urbain nous imposent de travailler plus particulièrement sur les ménages enquêtés au premier passage et qui ont été retrouvés au second passage.

Par ailleurs l'espace périphérique étudié est découpé suivant deux maillages administratifs : la Zone de dénombrement et le Secteur. Ce sont les découpages géographiques dont nous disposons pour aborder l'analyse de l'organisation spatiale des quartiers spontanés de la périphérie de Ouagadougou :

- la **Zone de dénombrement** est une unité spatiale mise en place pour les besoins du recensement de la population de 1985 ; elle a pour intérêt de constituer une unité géographique intermédiaire entre la Parcelle (unité de collecte) et le Secteur (unité politique d'aménagement de l'espace urbain).

- le **Secteur** est une division administrative regroupant les zones de dénombrement qui a été mise en place pour remplacer l'ancien découpage traditionnel en quartier. Il exprime une décision administrative mise en place par le gouvernement et ayant pour fonction de remplacer l'ancien découpage traditionnel en quartiers. Par ailleurs, le découpage en 16 secteurs de la périphérie est celui qui a servi de définition au périmètre de l'enquête sur "l'extension urbaine".

Par conséquent le système d'échelles que nous voulons étudier, est composé des éléments suivants :

- l'unité statistique de référence, le ménage. La population d'étude est le ménage;

- les deux groupes de variables observés aux niveaux du Ménage et de la Parcelle, soit 21 variables totalisant 74 modalités;

- les quatre niveaux d'échelles pour l'analyse de l'information :
  - le Ménage, le niveau le plus fin, qui fournit l'unité statistique de base et correspond à un niveau de définition des variables : les données analysées à ce niveau ne sont pas agrégées,

- la Parcelle qui est le premier niveau d'agrégation considéré et correspond également à l'autre niveau de définition des variables,
- le Secteur et la Zone de dénombrement qui sont des niveaux d'agrégation spatial pouvant être cartographiés et correspondant à des découpages administratifs.

S-Secteur	(16 secteurs)
Z-Zone de dénombrement	(188 zones)
P-Parcelle	(840 parcelles)
M-Ménage	(925 ménages)

Avant d'identifier les principaux paramètres sociologiques et fonciers de l'organisation spatiale de la périphérie de Ouagadougou et déterminer les niveaux d'analyse pertinents pour un phénomène donné, nous allons exposer la démarche d'analyse de ce système d'échelles.

4. 1. 2. PRESENTATION DE LA DEMARCHE D'ANALYSE

Il ne s'agit pas, au travers des analyses, d'affiner une connaissance sur Ouagadougou mais de dégager les grandes tendances de l'organisation sociale, résidentielle et spatiale de sa périphérie. C'est pourquoi nous envisageons d'une part d'examimer les plans factoriels principaux de chaque analyse (dont les aides à l'interprétation sont présentées dans [PIRON, 1990]) et d'autre part, de traiter en éléments actifs toutes les variables. En effet, puisque nous cherchons à **mettre en évidence les effets d'un changement d'échelle pour un même jeu de variables**, nous ne mettrons aucune modalité en supplémentaire même si parfois l'analyse le suggère.

Nous rappelons que l'étude du système d'échelles repose sur deux types d'analyses :

- **l'analyse des niveaux** : il s'agit de l'étude de la série de tableaux de correspondances, c'est-à-dire de l'analyse inter-classes pour la partition définie par un niveau. Elle donne l'organisation des données à ce niveau d'analyse et rend compte des éléments de regroupement ou de différenciation des classes déterminant ce niveau.

- **l'analyse du passage d'un niveau à un autre** : il s'agit de fournir les indicateurs de disparité à l'intérieur des classes définies par les niveaux d'agrégation (parcelles, zones de dénombrement, secteurs). Ces indicateurs sont donnés par des associations de variables. On envisage les passages successifs d'un niveau à un autre, c'est-à-dire on réalise l'analyse intra-classes d'une analyse inter-classes, à l'exception du passage concernant le niveau le plus fin, le Ménage, où nous avons une structure de base (disjonctive complète) ne pouvant être décomposée.

Elle donne l'information contenue au niveau inférieur qui est perdue dans le passage et rend compte de la variabilité interne des unités de ce niveau au niveau supérieur.

Ainsi, la démarche proposée pour l'analyse d'un système d'échelles est la suivante :

1- On réalise **une analyse au niveau le plus fin, le Ménage,** pour obtenir l'organisation des données du maximum d'information.

2- Puis, au niveau immédiatement supérieur, la Parcelle, on réalise une double analyse pour observer **le changement d'échelle du niveau Ménage au niveau Parcelle** : l'analyse du niveau, l'analyse du passage entre les deux niveaux.

On répète cette série de deux analyses aux niveaux supérieurs suivants. On s'intéresse alors :

3- **au changement d'échelle du niveau Parcelle au niveau Zone de dénombrement.**

4- **au changement d'échelle du niveau Zone de dénombrement au niveau Secteur.**

Le processus de l'analyse d'un système d'échelles revient à faire des "va et vient" d'une échelle à une autre pour comprendre ce qui se passe à toutes les étapes du système. Celles-ci sont schématisées de la manière suivante :

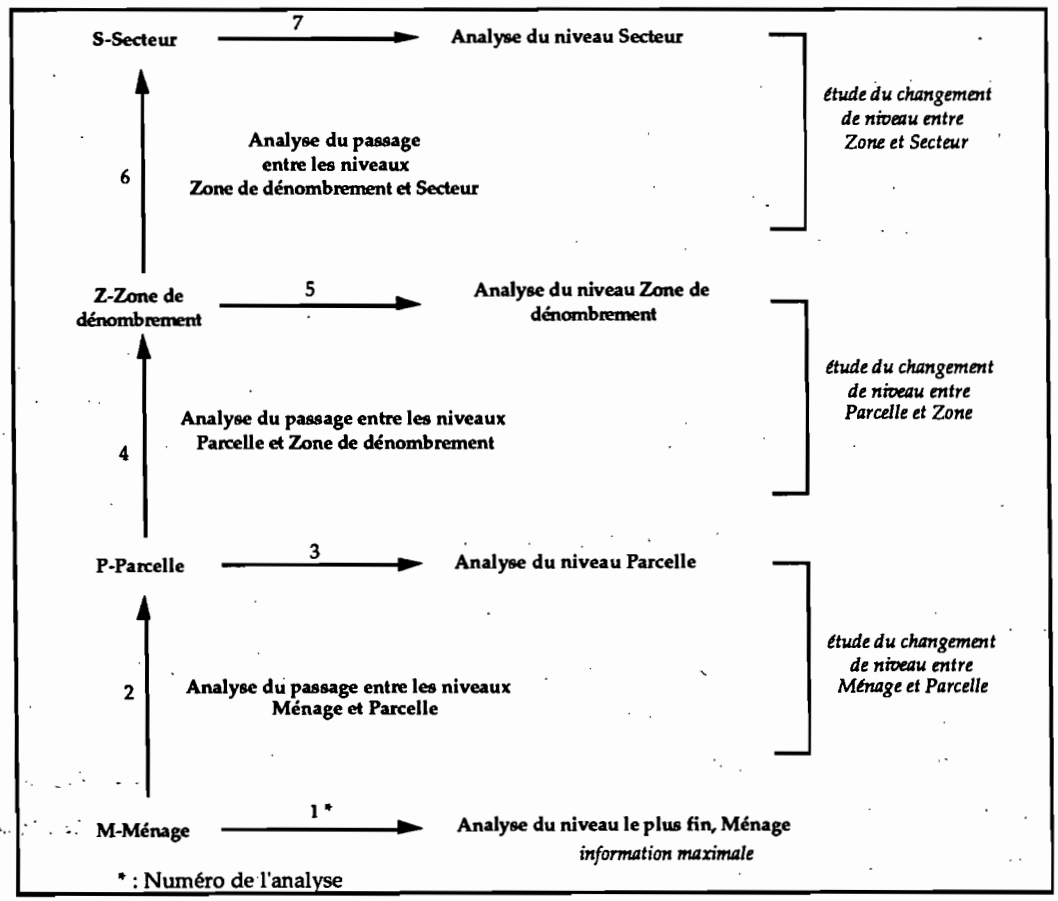


Figure n° 23 : *Les analyses du système d'échelles*

Par ailleurs, il semble particulièrement intéressant, compte tenu de la nature géographique des niveaux d'agrégation supérieurs, le Secteur et la Zone de dénombrement, de cartographier les résultats des analyses décrivant le changement d'échelle entre ces deux niveaux. L'intérêt est de déduire visuellement la position des unités spatiales sur le plan factoriel et ainsi localiser les phénomènes dégagés par ce même plan. Par conséquent, il s'agit de concevoir la légende de manière à traduire d'une part les oppositions que dégage chaque axe et d'autre part l'indépendance des deux axes.

L'ANALYSE DU SYSTEME D'ECHELLES

La méthode proposée doit permettre d'appréhender différents niveaux de collecte et de comprendre l'organisation d'un phénomène à différentes échelles. Dans cette application, on envisage un système d'échelles géographiques et on cherche à mettre en évidence les principaux facteurs de la structuration sociale et foncière de l'espace périphérique ouagalais.

4. 2. 1. ANALYSE AU NIVEAU LE PLUS FIN : LE MENAGE

Il s'agit de dégager l'organisation du maximum d'information à partir des variables descriptives de la Parcelle et du Ménage prises au niveau le plus fin : le Ménage. On considère l'analyse n°1 (fig. n°24) obtenue à partir du tableau disjonctif complet.

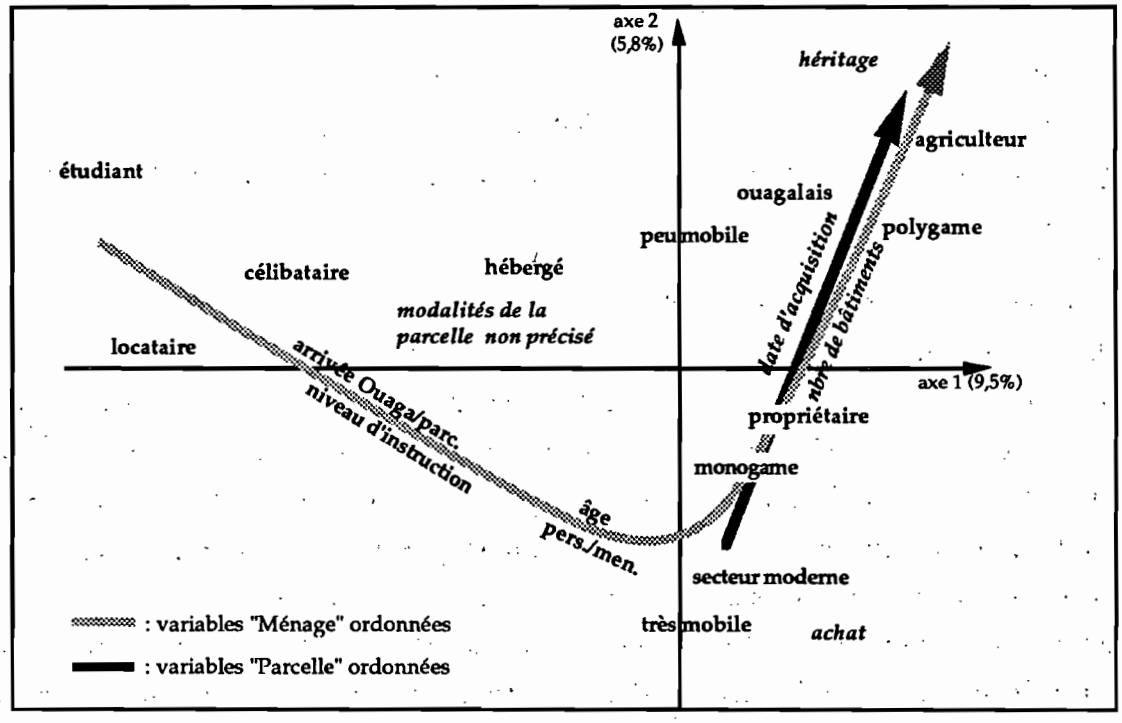


Figure n°24 : *Analyse n°1 du niveau Ménage*

La courbe de l'histogramme des valeurs propres indique une structure très marquée sur les deux premiers axes et des différenciations de ménages moins nettes ou très secondaires sur les axes

suivants. La valeur particulièrement élevée du premier facteur (comparativement aux autres) indique que celui-ci est déterminant dans l'organisation du nuage au niveau Ménage.

En ce qui concerne les variables relatives au ménage, on observe sur l'axe 1 une régularité dans l'ordre des coordonnées des variables ordonnées : "âge", "nombre de personnes dans le ménage", "dates d'arrivée à Ouagadougou", "date d'installation sur la parcelle" et "niveau d'instruction"; les valeurs élevées sont projetées sur la partie positive de l'axe et les valeurs faibles sur sa partie négative. Les valeurs intermédiaires sont positionnées autour de l'origine et ne définissent pas aussi bien cet axe. Les variables "statut matrimonial" et "statut d'occupation" contribuent également à la construction de **ce premier facteur** que l'on peut traduire par **le cycle de vie** en opposant :

- les valeurs faibles des variables ordonnées associées aux modalités "célibataire", "étudiant" mais aussi "locataire" caractérisant **le début du cycle** ;

- les valeurs élevées de ces variables associées aux modalités "polygame" et "propriétaire" définissant **la fin du cycle**.

Les valeurs intermédiaires des variables ordonnées apparaissent sur la partie négative de l'axe 2 et s'opposent aux valeurs extrêmes. **Le deuxième facteur**, au travers de ces variables, permet ainsi de nuancer l'information apportée par le premier axe et affine cette connaissance en faisant ressortir deux autres éléments de différenciation entre les ménages : **la mobilité intra-urbaine** (peu et très mobiles) **et le secteur d'activité** (agricole et moderne).

L'information apportée par ces deux axes se complètent et l'examen du plan factoriel principal met, par conséquent, en évidence un effet Guttman défini par la forme parabolique de la courbe reliant les modalités des variables ordonnées. Il apparait donc clairement que la principale différenciation entre les ménages est donnée par le cycle de vie.

La définition du premier facteur par toutes les modalités du statut d'occupation qui, en revanche ne construisent pas l'axe 2, montre que l'accès au logement est traduit par un effet d'âge : il existe une nette opposition entre les locataires (jeunes) et les propriétaires (âgés) ; les hébergés en revanche ne sont pas spécifiques d'une étape du cycle de vie (modalité projetée au barycentre des valeurs extrêmes des variables ordonnées sur l'axe 1 et opposée aux valeurs intermédiaires sur l'axe 2). De plus, cette variable "statut d'occupation" (contribuant uniquement à former à l'axe 1) et la variable "mobilité" (contribuant uniquement à construire l'axe 2) sont indépendantes sur ce plan factoriel. Quant à celle décrivant le "secteur d'activité", elle participe à la formation des deux axes. Les "étudiants"(sur l'axe 1), le "secteur moderne" et le "secteur agricole" (sur l'axe 2) marquent respectivement les étapes du début à la fin du cycle de vie. Par conséquent, il existe une relation marquée entre cette variable et la "mobilité" ; les chefs de ménage qui sont au début ou en fin du cycle de vie sont peu mobiles, et ceux qui correspondent aux étapes intermédiaires du cycle ont effectué

d'avantage de migrations. Cependant, la modalité "étudiant" est écartée de cette relation puisqu'elle intervient sur l'axe 1 : en effet, puisqu'ils sont jeunes, les étudiants ont effectué peu de migrations. Par ailleurs, le secteur d'activité est fortement en relation avec le "statut d'occupation ; les étudiants sont locataires ou hébergés, les chefs de ménage exerçant une activité dans le secteur moderne sont locataires ou propriétaires et ceux du secteur agricole sont propriétaires ou hébergés.

En ce qui concerne les variables descriptives de la parcelle,les modalités "non précisé" déterminent la partie négative de l'axe 1 et sont opposées aux autres modalités des variables du mode d'accès au sol (date et mode d'acquisition de la parcelle). **Le deuxième axe oppose, quant à lui, les modes d'accès traditionnels et anciens aux modes d'accès modernes et récents.**

L'examen de ce plan (schématisé par la figure n°30a) met en évidence les correspondances entre les variables descriptives du ménage et de la parcelle qui définissent trois types de chefs de ménage associés à un mode d'accès au sol et au logement :

- "**étudiant**": caractérisé par un chef de ménage jeune, célibataire, ayant un niveau d'étude supérieur et vivant seul. Ces ménages n'ont jamais migré dans la ville car leur arrivée à Ouagadougou et donc sur leur parcelle est récente. **Les renseignements sur les parcelles correspondantes n'ont pas été fournis.** Cela s'explique par le fait que ce type de ménage est locataire ou hébergé. Il n'est par conséquent pas toujours en mesure de donner les informations concernant la parcelle.

- "**ancien**": caractérisé par un chef de ménage âgé, veuf ou polygame, ayant plus de 10 personnes à charge et non scolarisé. Ces ménages, fortement représentés dans le secteur agricole et parmi les natifs de Ouagadougou, sont peu ou pas mobiles et sont installés à Ouagadougou depuis longtemps. Ils sont hébergés ou propriétaires dans quel cas ils sont **acquis leur parcelle d'une manière traditionnelle** et possèdent de nombreux bâtiments.

- "**moderne**": caractérisé par un chef de ménage monogame, âgé entre 30 et 40 ans, propriétaire ou locataire. La mobilité intra-urbaine, définie essentiellement par un déplacement du centre vers la périphérie, est nettement plus importante pour ces types de ménage qui résident principalement sur les parcelles d'**acquisition moderne et récente.**

4. 2. 2. CHANGEMENT D'ECHELLE DU NIVEAU MENAGE AU NIVEAU PARCELLE

4. 2. 2. 1. Niveau Parcelle

L'analyse du niveau Parcelle a une configuration quasiment semblable à celle réalisée au niveau Ménage puisque le nombre de ménages par parcelle est très faible (1,2 ménages par parcelle) : l'analyse à ce niveau d'agrégation, la Parcelle, conserve la stabilité de l'organisation originelle obtenue au niveau Ménage. De ce fait, le plan principal du niveau Parcelle n'est pas présenté ici dans le détail mais il est schématisé par la figure n°30c.

Comparativement à l'analyse du niveau le plus fin, les contributions des variables Ménage sont sensiblement inférieures et celles des variables Parcelle sont légèrement supérieures, notamment "le groupe polyrésidentiel" qui participe d'avantage à la construction du deuxième axe. Les ménages de même profil qui cohabitent entre eux sont soit des ménages du type "étudiant" soit des ménages du type "ancien". Ce dernier élément confirme la **maintenance de la société rurale ouagalaise dans ses formes traditionnelles.**

4. 2. 2. 2. Passage entre les niveaux Ménage et Parcelle

La variabilité interne des parcelles étant très faible, il existe une forte ressemblance entre l'analyse du niveau Ménage et celle du niveau Parcelle.

L'analyse du passage entre ces deux niveaux (fig. n°25) est cependant très intéressante car elle donne les profils de ménage différents qui cohabitent au sein d'une même parcelle : elle dégage les **éléments de variabilité interne des ménages à la parcelle.**

Elle exclue, en toute logique, d'une part les caractéristiques attachées à la parcelle, d'autre part, les ménages monorésidents et les groupes polyrésidentiels dont les ménages ont des profils parfaitement identiques. Elle met, par contre, en valeur les caractéristiques qui interviennent le plus dans la différenciation des ménages résidents sur une même parcelle.

Ce sont les mêmes variables et pratiquement les mêmes modalités de ces variables qui construisent les deux premiers facteurs. Le statut d'occupation, au travers des propriétaires et des hébergés, contribuent à la formation des deux premiers axes ce qui permet de définir sur la première bissectrice une **opposition entre les propriétaires et les hébergés.** On observe un effet Guttman pour les variables ordonnées (dates d'arrivée à Ouagadougou et sur la parcelle, âge) mais les valeurs extrêmes s'opposent d'une part aux valeurs intermédiaires sur l'axe 1 et d'autre part entre elles sur l'axe 2.

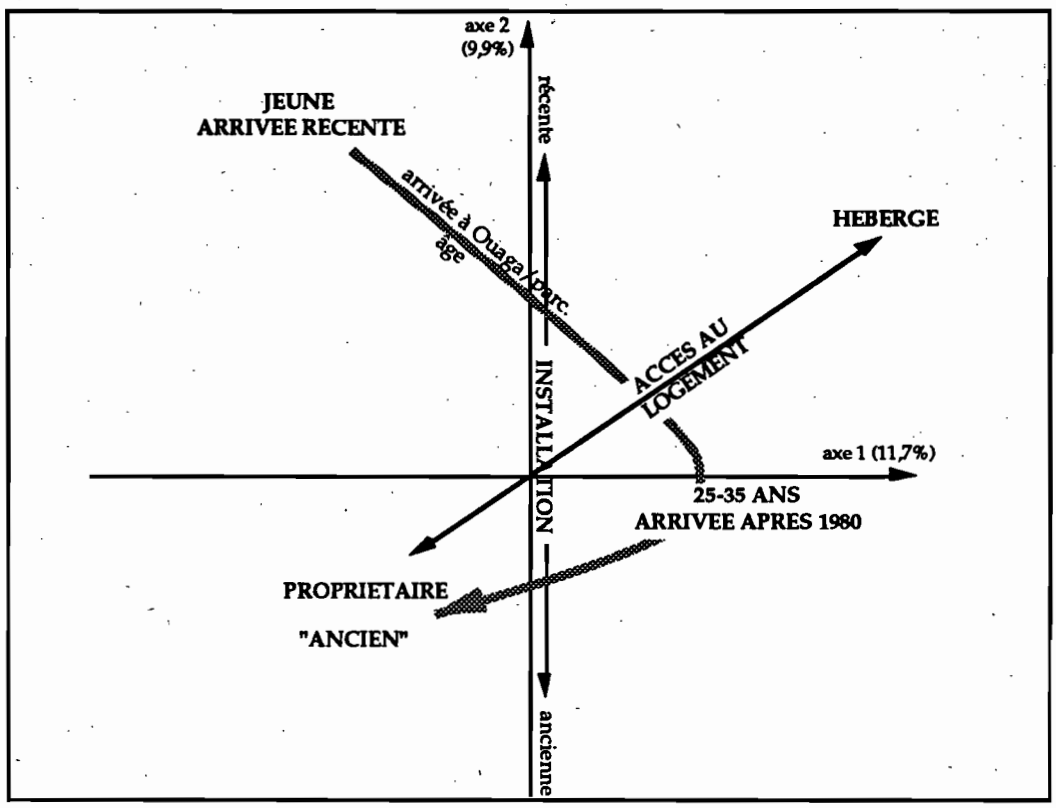


Figure n°25 : *Analyse n°2 du passage entre les niveaux Ménage et Parcelle*

Par conséquent, les principaux éléments de différenciation des ménages au sein d'une même parcelle sont essentiellement la date d'arrivée à Ouagadougou et la date d'installation sur la parcelle, puis le statut d'occupation et l'âge. On observe trois types de ménage :

- les "**anciens-propriétaires**", polygames, anciennement installés dans la ville et ayant résidé dans le centre ; dans une moindre mesure, les agriculteurs sont inclus dans ce type de ménage composé de 10 personnes ;
- les "**chefs de ménage arrivés après 1980**", monogames, 30 ans et non mobile, propriétaires ou hébergés ;
- les "**jeunes** (15-25 ans) **venant d'arriver à Ouagadougou**" (après 1983) et ayant moins de 3 personnes à charge.

Cette analyse fait ressortir essentiellement une forme de **cohabitation résidentielle** particulière : celle des **"propriétaires-hébergés"**. Ceci différencie Ouagadougou des autres grandes villes de l'Afrique de l'ouest où l'on rencontre surtout une cohabitation de "propriétaires-locataires".

4. 2. 3. CHANGEMENT D'ECHELLE DU NIVEAU PARCELLE AU NIVEAU ZONE DE DENOMBREMENT

4. 2. 3. 1. Passage entre les niveaux Parcelle et Zone de dénombrement

L'analyse n°4 du passage entre les niveaux Parcelle et Zone de dénombrement (fig. n°26) met en évidence les profils de parcelles distincts, définis par les caractéristiques socio-économiques des ménages y résidant et par les caractéristiques des parcelles, appartenant à une même zone de dénombrement.

Globalement, on retrouve des éléments de différenciation des parcelles au sein d'une même zone de dénombrement communs à l'analyse du niveau Parcelle, et donc à celle du niveau Ménage (structures des analyses n°1 et 4 sont sensiblement identiques (cf fig. n°24 et 26)). Ceci indique une grande diversité des parcelles mais aussi des ménages au sein d'une même zone de dénombrement : il existe une variabilité interne des zones de dénombrement non négligeable.

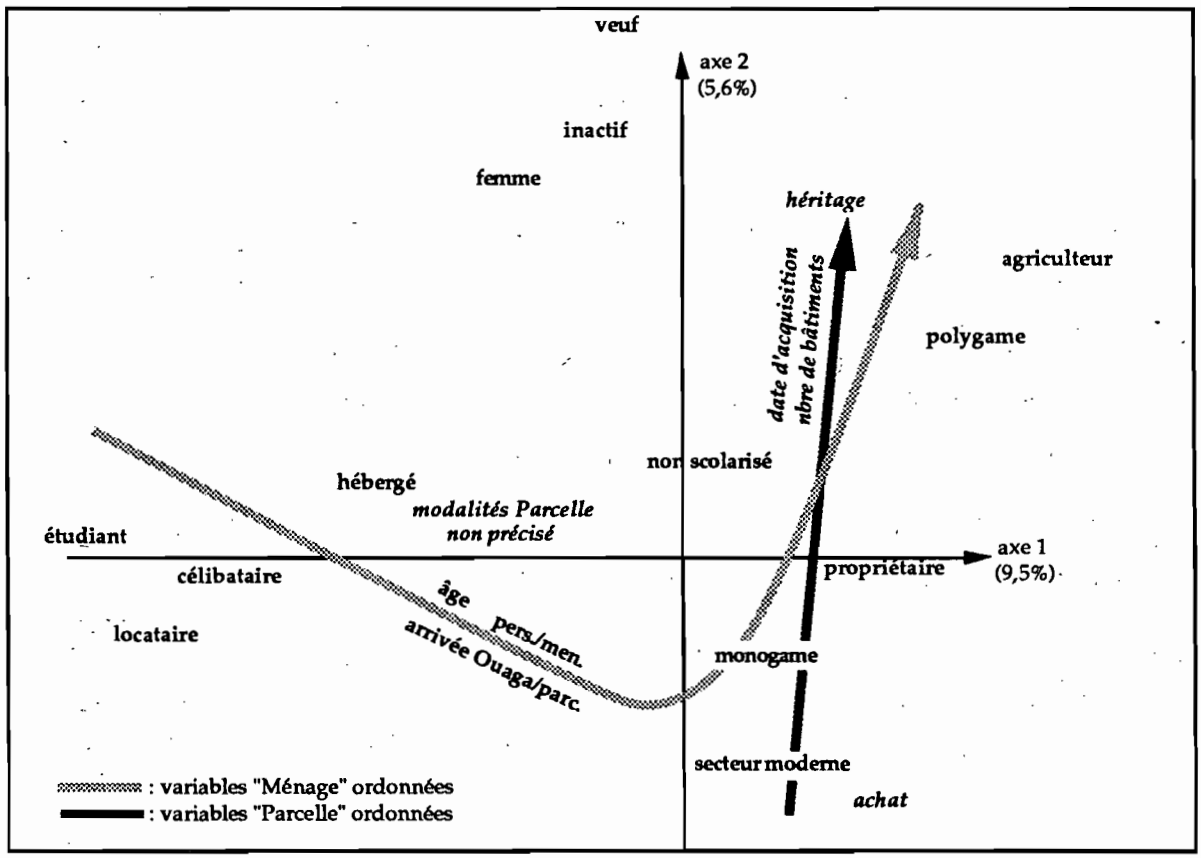


Figure n°26 : *Analyse n°4 du passage entre les niveaux Parcelle et Zone de dénombrement*

En effet **l'axe 1**, qui exprime une différenciation très nette (9,5%), **est marqué** par le phénomène de **cycle de vie** qui exclue cependant le "niveau d'instruction". Cette dernière variable apparait sur l'**axe 2** lequel **traduit** surtout **le secteur d'activité** en opposant le secteur agricole (partie positive) au secteur moderne (partie négative). Il oppose également les non scolarisés (partie positive) des scolarisés jusqu'aux études secondaires (partie négative).

A l'examen du plan factoriel principal, on retrouve l'effet Guttman très marqué et dont l'axe de symétrie de la parabole est la deuxième bissectrice. Il y a par rapport à l'analyse n°1 (et donc à l'analyse n°3) un décalage qui s'explique par la position des modalités "femme", "inactif" et "veuf" sur l'axe de la parabole. Ce groupe de variables exprime une indépendance avec le phénomène "cycle de vie".

Les variables descriptives de la parcelle qui contribuent à la formation des deux axes sont les mêmes que l'analyse du niveau Parcelle (ou Ménage).

Quatre types de parcelles et de ménages y résidant (schématisés dans la figure n°30d) se distinguent et interviennent comme éléments d'hétérogénéité pour les zones de dénombrement :

- le type "**élève-étudiant**": il s'agit d'étudiants mais qui ne sont pas caractérisés par un niveau d'étude supérieur. De plus, ce type de ménage se définit comme étant locataire ou hébergé, récemment arrivé à Ouagadougou et récemment installé sur sa parcelle. Les renseignements concernant la parcelle n'ont pas été donnés.

- le type " **moderne**", où l'on dissocie :
  - les monogames ayant entre 6 et 9 personnes à charge et ayant acheté leur parcelle entre 1975 et 1980,
  - et dans une moindre mesure, les non mossi et locataires qui se rapprochent du type "étudiant".

- le type "**ancien-agriculteur**": caractérisé par un chef de ménage polygame, âgé, ayant à charge plus de 10 personnes et exerçant une activité dans le secteur agricole. Propriétaire de sa parcelle par héritage ou par attribution coutumière, il y est installé depuis longtemps.

- le type "**veuf**": caractérisé par un chef de ménage féminin et sans activité professionnelle. Ce type de ménage est au barycentre des types "étudiant" et "ancien-agriculteur" et est situé sur l'axe de la parabole décrivant le cycle de vie. Il n'est pas présent dans l'analyse du plan principal du niveau Parcelle mais il apparait sur l'axe 3 de cette analyse.

Cette analyse du passage entre les niveaux Parcelle et Zone de dénombrement met en évidence les types de parcelles en relation avec des types de ménages qui sont dispersés au sein des zones de

dénombrement. Ces types sont essentiellement définis par les variables socio-économiques. Par ailleurs, le nuage des individus-parcelles est bien réparti sur tout le plan et s'allonge d'avantage sur la partie positive de l'axe 2 (le type "veuf" et dans une moindre mesure le type "étudiant" sont plus marginaux ce qui n'est pas le cas pour les types "ancien" et "moderne"). Ceci permet de conclure que **s'il existe des regroupements significatifs de ménages de même type socio-économique au sein d'une même parcelle** (rappelons que la plupart des parcelles abritent une seul ménage), **il y a une plus grande diversité de ceux-ci au sein d'une même zone de dénombrement.**

4. 2. 3. 2. Niveau Zone de dénombrement

L'analyse n°5 du niveau Zone de dénombrement (fig. n°27) met en évidence des types de zones de déombrement ayant des profils globalement identiques quant aux types de parcelles et de ménages les composant.

Cette analyse présente un nuage structuré globalement de la même manière que ceux des analyses précédentes. Mais ce ne sont plus les mêmes variables qui contribuent à la définition des axes : il s'agit de variables descriptives de la parcelle et surtout de la situation de la parcelle par rapport aux opérations d'aménagement ce qui s'explique facilement puisque ces variables ont une masse plus importante. De ce fait **l'axe 1 met en évidence le phénomène de restructuration** en opposant les terrains qui sont en pleine restructuration (en cours de lotissement ou récemment loti sur la partie positive) de ceux qui ne le sont pas (anciennement loti ou non loti sur la partie négative). **L'axe 2** oppose ce qui est loti ou en cours de lotissement (partie positive) de ce qui est non loti ou récemment loti (partie négative). Il **traduit** également **le mode d'acquisition de la parcelle** en dissociant les modes d'acquisition traditionnels et de dates anciennes (partie positive) des modes d'acquisition modernes et de dates récentes (partie négative).

Par conséquent, les variables "Ménage" contribuent beaucoup moins à la construction des facteurs et l'effet Guttman très présent sur les autres analyses existe mais n'explique pas la formation de ce plan. La parabole est uniquement définie par les variables "dates d'arrivée à Ouagadougou et sur la parcelle".

Ce niveau d'analyse met en évidence **une relation entre les opérations d'aménagement** (restructuration ou non), **les modes d'acquisition** (traditionnel ou moderne) **et quelques types de ménages marginaux** qui se regroupent au sein d'une même zone de dénombrement : "étudiant-locataire", "ancien-agriculteur", "ancien-ouagalais".

On note que la variable "groupe polyrésidentiel", contribue largement à la formation de l'axe 2 et sa position au barycentre des types "étudiant-locataire" et "ancien-ouagalais" permet de déduire que ces types de ménages constituent chacun des groupes polyrésidentiels homogènes. Ceux-ci ne peuvent

apparaître dans l'analyse n°2 du passage entre les niveaux Ménage et Parcelle (cf fig. n°25) qui, rappelons-le, met en évidence une cohabitation entre ménages de profils différents. C'est le cas pour le type "ancien-agriculteur" qui, nous avons constaté, cohabitent avec des types de ménages différents mais pas entre eux. Il y a indépendance (orthogonalité) entre les "ancien-agriculteur" (sur l'axe 1) et la variable "polyrésidence" (sur l'axe 2) au niveau d'analyse de la Zone de dénombrement.

La nature géographique du niveau Zone de dénombrement permet de cartographier les résultats de cette analyse (fig. n°27). Les grandes tendances de l'organisation de l'espace propre à l'échelle de la Zone de dénombrement (schématisées par la figure n°30e) sont :

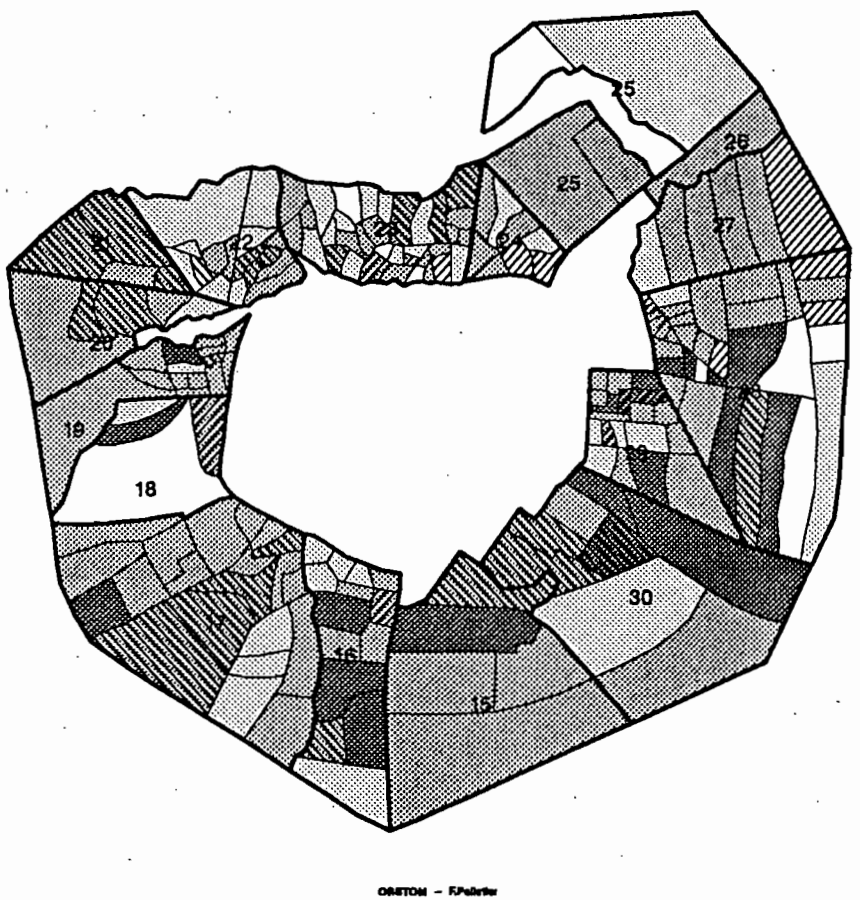
- au sud et à l'est de Ouagadougou, les zones regroupant les **ménages récemment arrivés** dans la ville et sur leur parcelle **en terrain loti ou non loti** du type célibataire, non mossi, de niveau d'instruction élevé, d'origine urbaine et récemment arrivés sur leur parcelle" (partie négative de l'axe 1).

- au nord de la ville, sur les terrains en pleine restructuration (en cours de lotissement ou récemment lotis), les zones surreprésentées par les ménages anciennement installés à Ouagadougou et sur leur parcelle dont **le mode d'acquisition est traditionnel** (parties positives des axes 1 et 2).

- dispersées sur l'ensemble de l'espace périphérique et en nombre important, les zones qui ne regroupent pas de types de ménages bien identifiés et qui se caractérisent d'avantage par un type d'accès au sol : **mode d'acquisition moderne et récent**. Les parcelles correspondantes se localisent principalement en terrain non loti ou récemment loti (partie négative de l'axe 2). On relève l'existence significative d'acquisitions récentes en zone non lotie.

Le type "étudiant-locataire", contribuant largement à la partie négative de l'axe 1 et positive de l'axe 2, est un phénomène très marginal et est très concentré essentiellement dans trois zones de l'est de la ville et proches de l'université. En revanche, la forte concentration du nuage des zones de dénombrement dans les trois autres quadrants indique que **l'espace appréhendé au travers des zones de dénombrement se différencie peu en fonction des types de ménages et d'avantage selon un état et un mode d'acquisition de la parcelle**.

Certains secteurs sont composés de zones très différenciées (secteurs 16, 19, 28, 29). Les éléments de différenciation entre les zones d'un même secteur sont mis en évidence par l'analyse du passage entre les niveaux Parcelle et Zone de dénombrement.



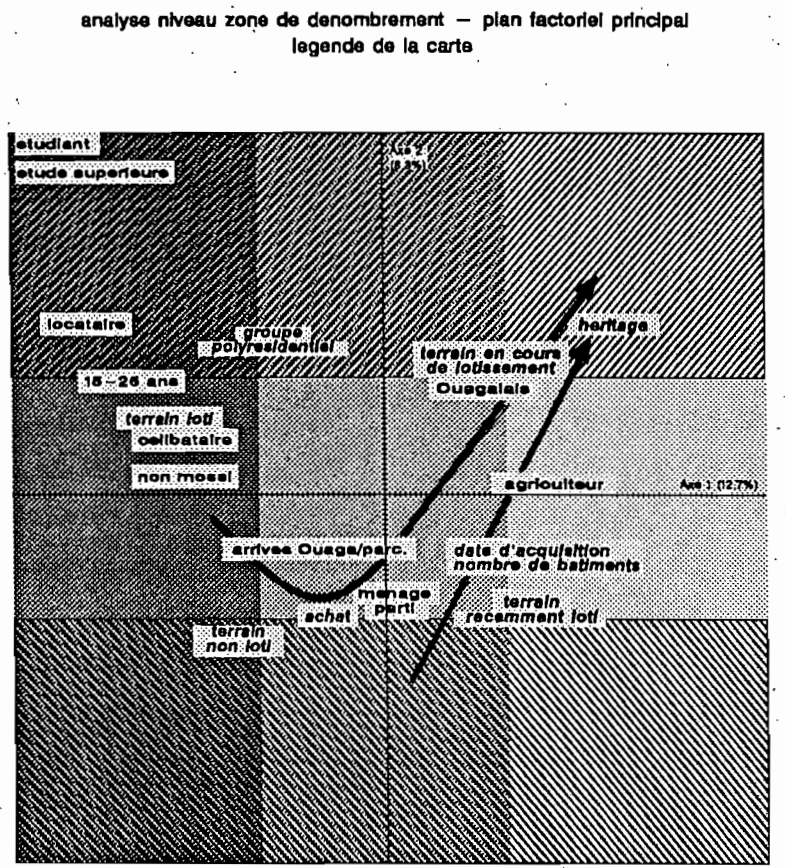


Figure n°27 : *Analyse n°5 du niveau Zone de dénombrement*

4. 2. 3. 3. Conclusion

Les deux analyses qui composent l'étude du changement d'échelle du niveau Parcelle au niveau Zone de dénombrement, mettent en évidence : au niveau Zone, une relation entre les types de parcelles décrites par ses propres variables et celles des ménages y résidant et les principales opérations de lotissement qui caractérisent ces zones ; dans le passage entre les deux niveaux, les types de ménages qui interviennent comme éléments de différenciation interne pour les zones de dénombrement.

Dans l'ensemble, il n'y a pas de relation entre les types de ménages définis dans les analyses des niveaux Ménage et Parcelle, et la nature des opérations de lotissement correspondant à la zone de dénombrement. On peut, cependant, retrouver des types de ménages (type "agriculteurs") communs aux deux analyses de l'étude de ce changement d'échelle. Cela signifie qu'ils peuvent être à la fois éléments de dispersion pour certaines zones de dénombrement et éléments de caractérisation pour d'autres zones.

4. 2. 4. CHANGEMENT D'ECHELLE DU NIVEAU ZONE DE DENOMBREMENT AU NIVEAU SECTEUR

4. 2. 4. 1. Passage entre les niveaux Zone de dénombrement et Secteur

L'analyse n°6 du passage entre les niveaux Zone de dénombrement et Secteur (fig. n°28) fait ressortir les différents profils de zones de dénombrement qui appartiennent à un même secteur.

La structure du nuage est globalement la même que celle du nuage pris au niveau Zone de dénombrement bien que ce ne soit pas tout à fait les mêmes variables qui contribuent à la construction des facteurs. En effet, la variable "opérations de lotissement" n'intervient pas pour différencier un même secteur, à l'exception de la modalité "anciennement loti" qui est en relation avec des variables descriptives du ménage (type "étudiant"). Ce sont essentiellement les variables spécifiques de la parcelle qui déterminent les axes mais certains éléments de différenciation des zones au sein d'un même secteur sont fournis également par les combinaisons de variables descriptives du ménage. Ainsi on retrouve la parabole significative d'un effet Guttman qui retrace le phénomène du "cycle de vie", masi elle est essentiellement construite par les variables "dates d'arrivée à Ouagadougou" et "date d'installation sur la parcelle". Les valeurs intermédiaires des autres variables ordonnées ne contribuent pas à la formation du plan principal.

L'axe 1 traduit par conséquent **le cycle de vie** en mettant surtout en évidence les étapes extrêmes (début (partie négative) et fin (partie positive)). **L'axe 2 définit le mode d'acquisition de la parcelle** en opposant les modes modernes sur sa partie négative aux modes traditionnels sur sa partie positive.

La carte du passage entre les niveaux Zone de dénombrement et Secteur permet de localiser les zones qui sont éléments d'homogénéité au niveau Zone de dénombrement et éléments de dispersion au sein des secteurs (fig. n°28 et figure n°30f) :

- le type "**étudiant-locataire**" constituant des groupes polyrésidentiels, localisé dans les terrains lotis avant le premier passage. Il s'agit ici d'un groupe marginal très localisé sur les zones proches de l'université.

- le type "**ancien**", ayant hérité de sa parcelle et possédant plusieurs bâtiments ce qui correspond à un **mode d'attribution traditionnelle et ancienne.** Les zones surreprésentées par ce type sont bien réparties sur l'espace périphérique.

- le type "**jeune**", célibataires, non mossi, récemment installés dans la ville, dans les terrains non lotis. Les zones spécifiques de ce type sont localisées sur le sud et l'est de la ville.

- le type de ménage résidant sur des parcelles "**d'attribution moderne et récente**" est caractéristique de zones de dénombrement dispersées sur l'ensemble de la périphérie.

Dans l'ensemble, **la variabilité interne des secteurs est importante : les secteurs sont constitués de zones fortement différenciées entre elles** (éloignées de l'origine du plan) **quant aux phénomènes dégagés par le plan factoriel principal** c'est-à-dire **la relation entre le cycle de vie et le mode d'acquisition de la parcelle.** Les secteurs 16, 19, 28, 29, sont constitués de zones très diversifiées quant aux structures déterminées par ce plan et les zones éliminées dans le passage vers le niveau supérieur, le Secteur, sont caractérisées pour l'essentiel par le cycle de vie et par le mode d'acquisition de la parcelle.

4. 2. 4. 2. Niveau Secteur

L'analyse n°7 du niveau Secteur (fig. n°29) présente une configuration très différente des autres analyses : plus de la moitié de l'inertie du premier facteur provient de la variable "**opérations de lotissement**" qui contribue également au quart de la formation du second facteur. La mise en élément supplémentaire de cette variable ou l'étude des facteurs de rangs plus élevés pourrait permettre de dégager les autres structures qui sont ici masquées par cette variable. Nous nous en tiendrons malgré

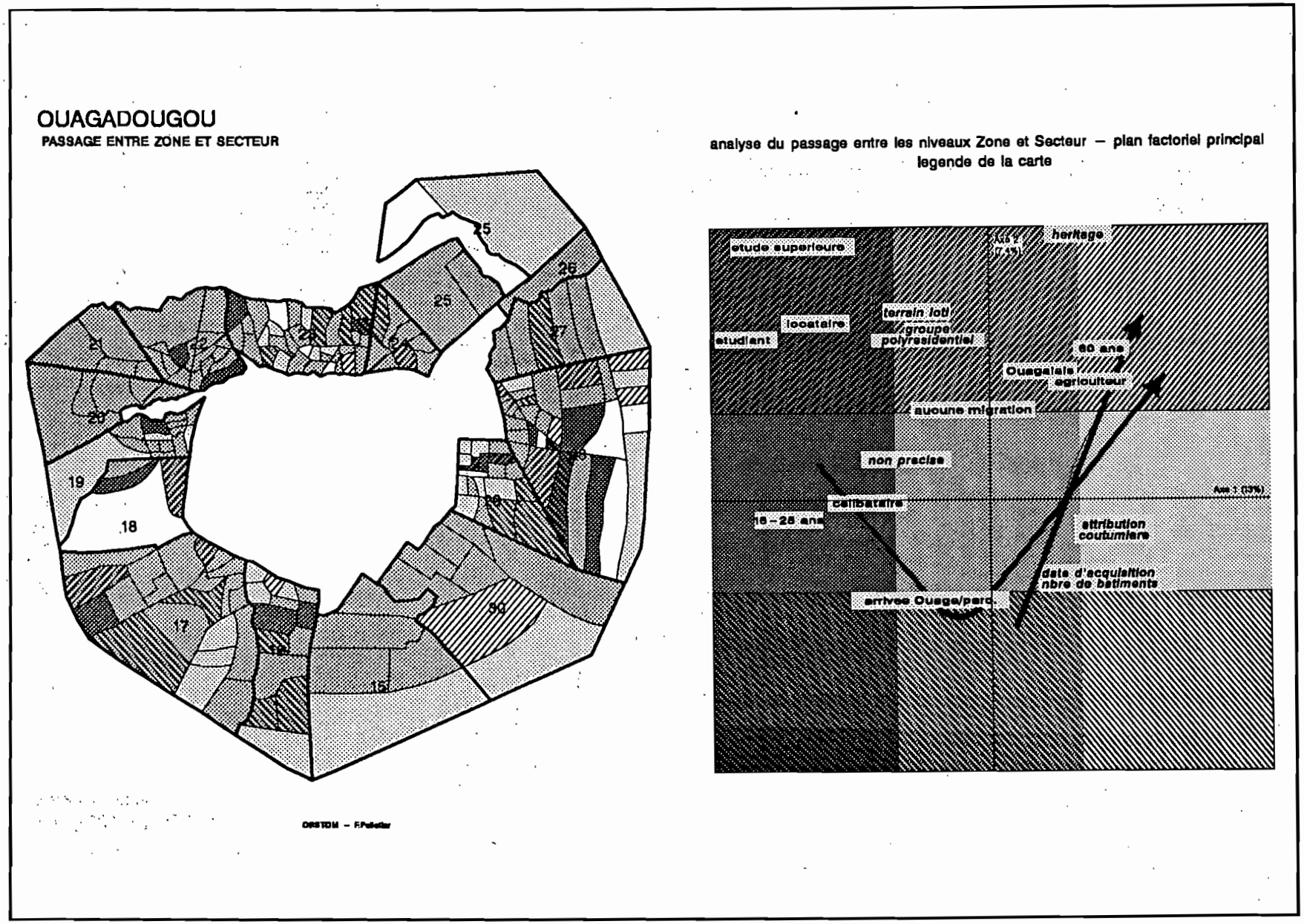


Figure n°28 : *Passage entre les niveaux Zone et Secteur*

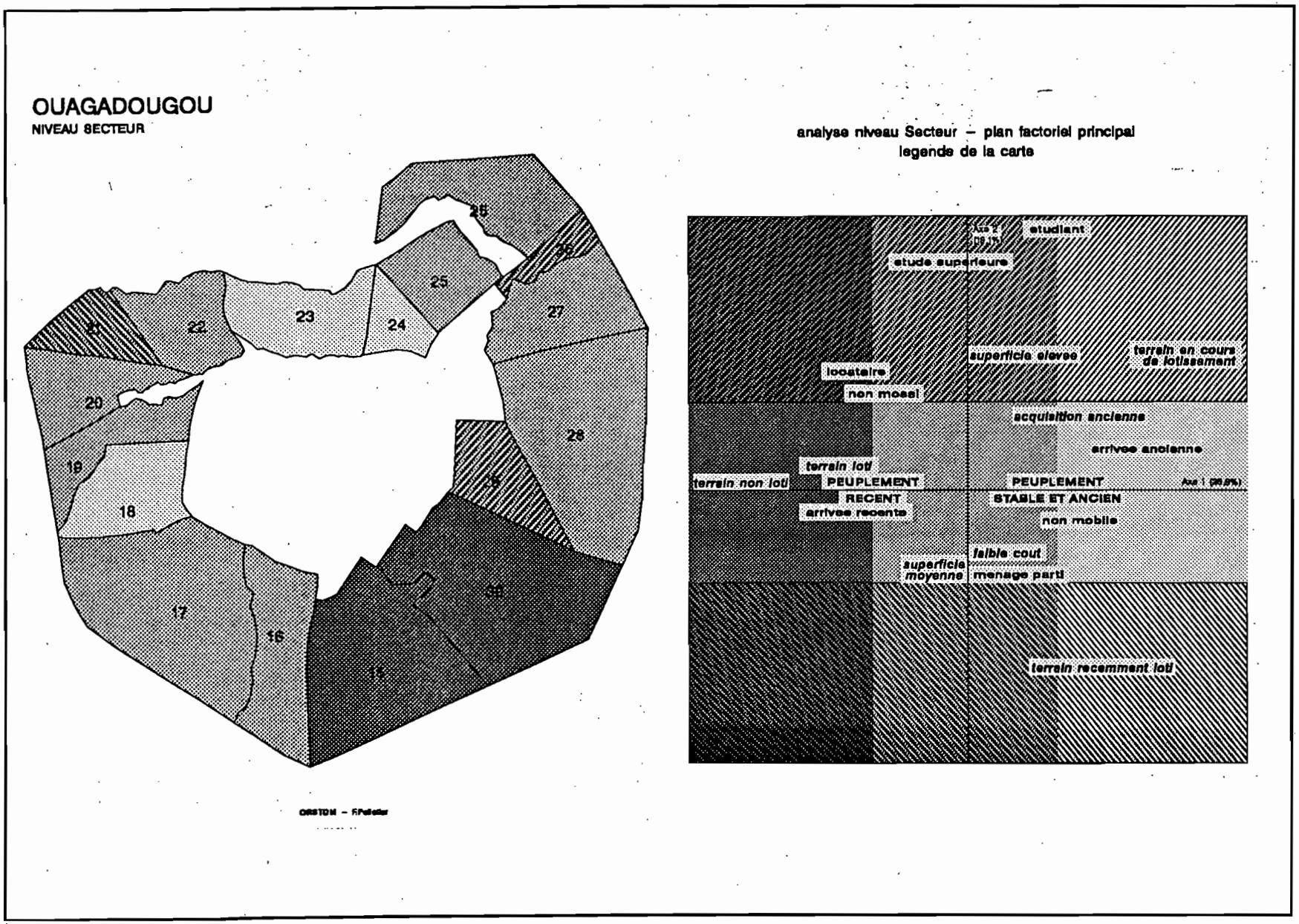


Figure n°29 : *Niveau Secteur*

tout à cette analyse d'une part pour les raisons évoquées en 4.1.2., et d'autre part parce que cette variable est centrale pour l'étude.

Par conséquent, **les deux axes traduisent le phénomène de restructuration** de l'espace périphérique:

- le premier axe oppose, comme celui de l'analyse du niveau Zone de dénombrement, les terrains qui sont en pleine **restructuration** (en cours ou récemment loti) de ceux qui ne le sont pas (anciennement ou non loti) et dans une moindre mesure les capacités des secteurs à accueillir ou à retenir leur population (mobilité intra-urbaine et date d'arrivée sur la parcelle).
  Les secteurs sur la partie négative sont surreprésentés par **des terrains anciennement lotis ou non lotis** et sont caractérisés par une **population récemment arrivée** sur leur parcelle.
  Les secteurs projetés sur la partie positive de l'axe sont **en pleine restructuration** et se distinguent par un **peuplement ancien et stable.**

- Le deuxième axe précise cette **restructuration** d'une part en dissociant les terrains qui sont en cours de lotissement de ceux qui viennent juste d'être lotis et d'autre part en dégageant **les caractéristiques morphologiques de l'habitat** (superficie et densité du bâti).
  Ainsi, sur la partie positive de l'axe, nous identifions les secteurs qui se caractérisent par des parcelles de superficie importante dont le mode d'acquisition est traditionnel et ancien, localisées sur des terrains en cours de lotissement. On trouve les types de ménages marginaux composés d'étudiants, de non mossi et les locataires.
  Sur la partie négative, sont projetés les secteurs surreprésentés par des parcelles récemment loties ayant un seul bâtiment et dont la superficie correspond aux normes des lotissements. Les **ménages partis** après le premier passage de l'enquête ont subi les opérations de déguerpissement de la politique d'aménagement.

Les nuages des variables et celui des secteurs sont très concentrés autour de l'origine du plan. Cela signifie que les secteurs sont proches du profil moyen désignant l'ensemble de la périphérie et sont peu différenciés entre eux. Si l'on poursuit l'examen des axes, l'espace défini par les secteurs, se différencie, en plus des opérations de lotissement, par l'origine du chef de ménage, son niveau d'instruction contribuant essentiellement à déterminer les axes plus lointains. La carte au niveau Secteur (fig. n°29) permet de visualiser les principales structures dégagées à ce niveau d'analyse et de dégager les types de secteurs suivants :

- les secteurs d'accueil (secteurs 15-30), surreprésentés par les terrains lotis ou non lotis et définis comme lieux de peuplement récent et de faible rétention (grande mobilité). Leur population est plutôt d'origine urbaine, non mossi et locataire.

- le secteur 29 (secteur contigu à l'université) est caractérisé à la fois par les zones en cours de lotissement et par la présence des "étudiants-locataires" et les non mossi qui contribuent à constituer des groupes polyrésidentiels.
- les secteurs de forte rétention (secteurs 23-24) caractérisés par des zones en pleine restructuration et définis comme lieux de peuplement ancien et stable.
- les secteurs surreprésentés par des terrains récemment lotis et où l'on compte une forte proportion de ménages partis entre les deux passages d'enquête (secteur 21 et dans une moindre mesure les secteurs 17 et 22)).

Globalement, les zones de dénombrement et les secteurs du nord, nord-ouest, se distinguent par des ménages anciennement arrivés à Ouagadougou et installés sur leur parcelle. A l'inverse, les zones et les secteurs du sud, sud-est, sont caractérisées par les ménages récemment arrivés à Ouagadougou.

4. 2. 4. 3. Conclusion

L'information obtenue au niveau Zone de dénombrement a été transformée. L'étude simultanée des trois cartes illustre bien la schématisation de la décomposition de l'inertie représentée par la figure n°20 (§3.3.1. et 3.3.2.). L'examen des deux premières cartes montre le recentrage des structures autour des centres de gravité de chaque secteur. Celui-ci est peu marqué pour les secteurs dont les zones sont très diversifiées (secteurs 16, 18), ce qui indique une forte variabilité à l'intérieur de ces secteurs. Il est, par contre, plus prononcé pour les secteurs homogènes (secteurs 21-26) et le secteur 26, qui ne contient qu'une seule classe, est un exemple significatif de faible variabilité (nulle). Dans le premier cas (hétérogénéité interne des secteurs), cela se traduit par des secteurs concentrés autour de l'origine au niveau Secteur (fig. n°29) et dans le deuxième cas (homogénéité des secteurs), par une forte proportion de zones au sein d'un même secteur projetées au centre du plan (fig. n°28).

4. 2. 5. SYNTHESE

La série d'analyses des passages entre les niveaux permet d'aborder la structuration de l'espace urbain, en dégageant les **indicateurs de disparité** à l'intérieur des classes définies par chaque niveau, et la série d'analyses des niveaux d'agrégation étudie cette structuration en dégageant les **spécificités des classes de chaque niveau**. Ainsi il est possible de définir l'organisation dominante à chaque échelle et par des "aller-retour" de comprendre les éléments de la structuration de l'espace urbain défini au travers des parcelles, des zones de dénombrement et des secteurs. Les principales

structures de cette série d'analyses sont schématisées dans la figure n°30 et, d'un point de vue méthodologique, peuvent être mises en parallèle avec les figures n°20 et 21 (cf § 3.3.).

Ce système d'échelles est intéressant car il permet de rendre compte des trois cas de figures possibles pour identifier l'homogénéité des niveaux (cf § 3.3.3.) : homogénéité très marquée pour certains niveaux, hétérogénéité très marquée pour d'autres niveaux ou modification progressive des phénomènes d'un niveau à un autre.

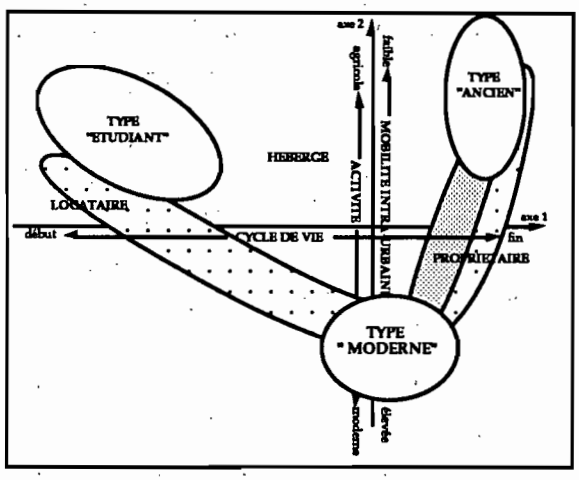
Globalement, les structures des nuages sont sensiblement équivalentes à l'exception des deux analyses qui définissent une organisation des données très différente : celle du passage entre les niveaux Ménage et Parcelle et celle du niveau Secteur. Cela s'interprète de la manière suivante :

- l'analyse du passage entre les niveaux Ménage et Parcelle (fig. n°30b) a une configuration très différente de l'analyse du niveau Ménage (fig. n°30a) : l'analyse au niveau d'agrégation Parcelle (fig. n°30c) conserve la stabilité de l'organisation originelle obtenue au niveau Ménage et ce niveau est constitué de parcelles très homogènes (ce qui est évident puisque la majorité des parcelles ne compte qu'un ménage).

- l'analyse au niveau Secteur (fig. n°30g) ne se structure pas comme celle du niveau immédiatement inférieur, la Zone de dénombrement (cf fig. n°30f). L'analyse du passage entre les niveaux Zone et Secteur et celle du niveau Zone de dénombrement rend compte globalement d'une même forme de nuage. Les secteurs sont globalement hétérogènes c'est-à-dire sont composés de zones de dénombrement de profils très différenciés donnés par le plan factoriel principal.

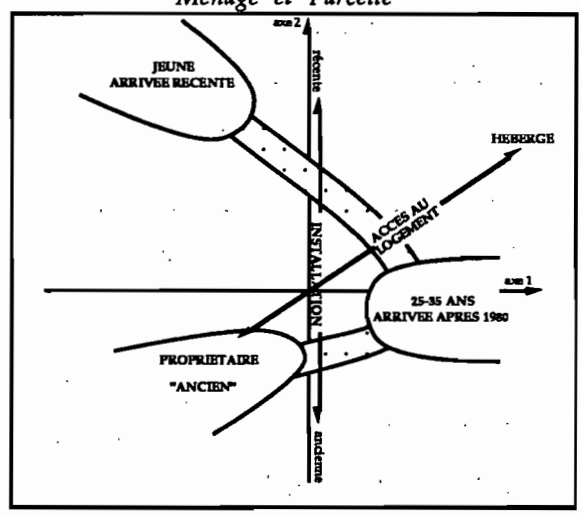
- à l'examen des analyses des niveaux Ménage (fig. n°30a), Parcelle (fig. n°30c), puis Zone (fig. n°30e) et des passages entre Parcelle et Zone (fig. n°30d) et entre Zone et Secteur (fig. n°30f), nous retrouvons une même forme de nuage décrivant le cycle de vie en relation avec le mode d'acquisition de la parcelle bien que ce ne soient pas les mêmes éléments qui contribuent à chaque fois à la construction des facteurs. Cela signifie qu'entre les niveaux Parcelle et Zone de dénombrement, les phénomènes qui reposent essentiellement sur une relation entre les structures résidentielles et socio-économiques, se modifient progressivement sur l'espace.
  La structure identique de ces nuages et la contribution importante de la variable "opérations de lotissement", au niveau Zone, (qui masque le rôle des variables ménages interviennant dans les analyses des niveaux plus fins), permettent de conclure qu'il **n'y a pas de relation entre les opérations d'aménagement et une organisation sociale sur la périphérie** de Ouagadougou : autrement dit, à la date de l'enquête (1987), les opérations d'aménagement n'ont pas encore généré de processus de ségrégation sociale.

fig. n°30a : *Niveau Ménage*



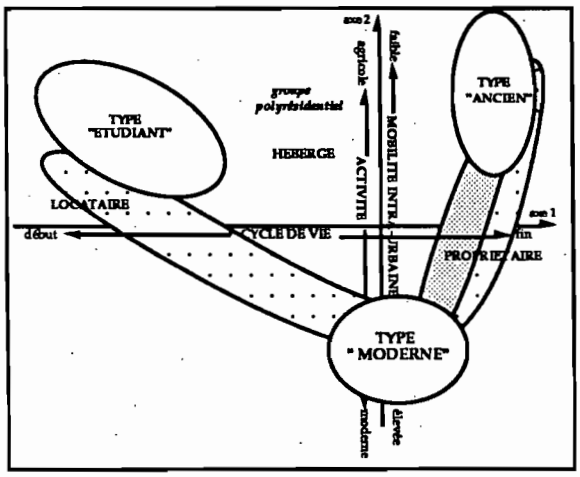
Mise en évidence de la relation [1] entre les structures résidentielles et socio-économiques. **(cycle de vie, mobilité et activité)**

fig. n°30b : *Passage entre les niveaux Ménage et Parcelle*



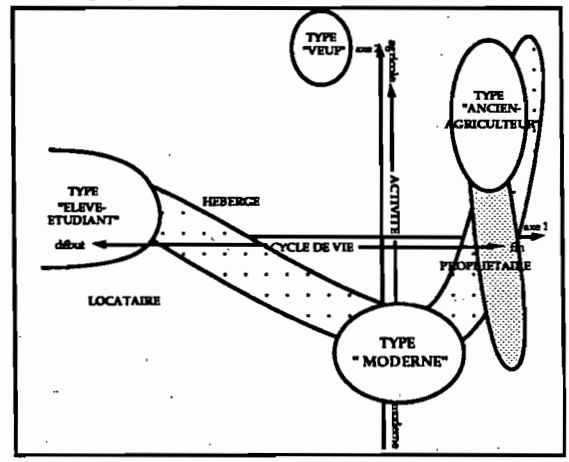
Perte sur la relation [1] et mise en évidence d'une cohabitation Propriétaire/hébergé. (relation [2] : **accès au logement**)

fig. n°30c : *Niveau Parcelle*



Permanence de la relation [1]. **(cycle de vie, mobilité et activité)**

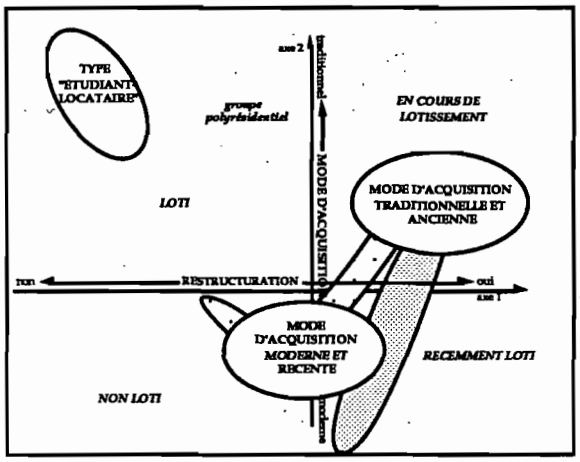
fig. n°30d : *Passage entre les niveaux Parcelle et Zone de dénombrement*



Perte de la mobilité sur la relation [1] : Mise en évidence des principales structures socio-économiques. **(cycle de vie et activité)**

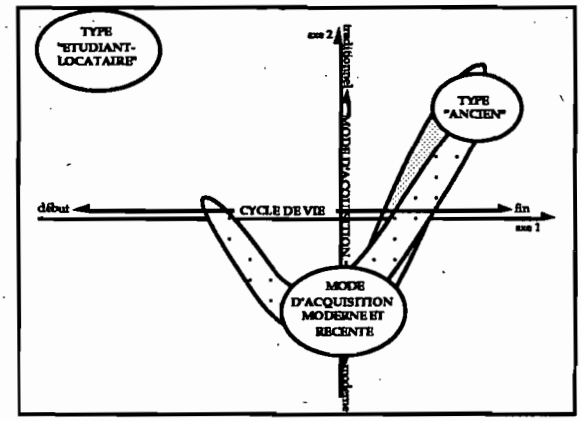
Figure n°30 (début): *Les analyses du système d'échelles*

fig. n°30e : *Niveau Zone de dénombrement*



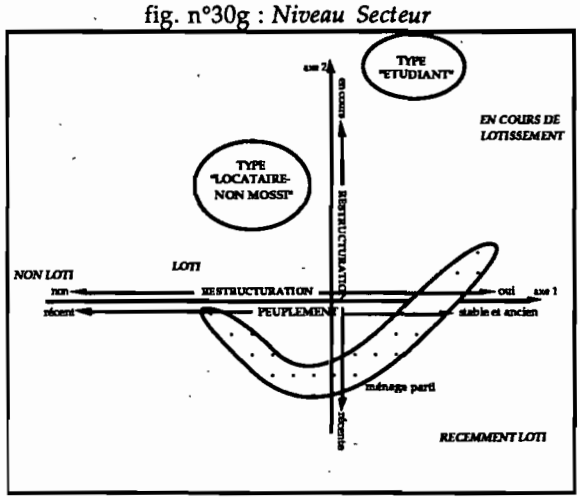
Mise en évidence de la relation [3] entre les opérations d'aménagement, les structures résidentielles et certaines structures socio-économiques.
(restructuration et mode d'acquisition de la parcelle)

fig. n°30f : *Passage entre les niveaux Zone de dénombrement et Secteur*



Permanence des relations [1] et [3]
Mise en évidence des principales structures résidentielles et socio-économiques.
(cycle de vie et mode d'acquisition de la parcelle)

fig. n°30g : *Niveau Secteur*



Mise en évidence de la relation [4] entre les opérations d'aménagement et l'intégration urbaine.
(Restructuration et peuplement)

Figure n°30 (fin) : *Les analyses du système d'échelles*

Cette absence de relation apparait au niveau de la Zone de dénombrement et se confirme au niveau du Secteur dont l'analyse n'identifie pas une correspondance entre une organisation sociale et une organisation spatiale. Seuls certains types de ménages très marginaux tendent à se regrouper géographiquement aux deux échelles géographiques. D'une manière plus globale, les quelques regroupements sociologiques sont définis surtout par la date d'installation à Ouagadougou et sur la parcelle. La périphérie se dichotomise : au nord les quartiers anciens et au sud les quartiers de peuplement plus récent.

L'INFLUENCE DU SYSTEME D'ECHELLES

Nous nous sommes intéressés, jusqu'à présent, aux facteurs qui caractérisaient les étapes dans le changement d'échelle. On cherche, maintenant, à étudier l'influence de l'échelle sur une variable, en déterminant le mode et l'intensité d'intervention de chaque variable, niveau par niveau.

4. 3. 1. COMPARAISON ENTRE LES TAUX D'INERTIE

Dans un but expérimental pour observer l'influence de l'échelle, nous introduisons deux systèmes d'échelles fictifs construits sur le même nuage de points de l'étude (c'est-à-dire définis par des unités statistiques, des variables et une population semblables). Les découpages en classes de ces systèmes d'échelles correspondant à chaque niveau sont identiques entre eux et avec celui de l'étude. Ils doivent également respecter la structure hiérarchique de l'information. Pour cela, le niveau Parcelle est conservé, les niveaux Zone de dénombrement et Secteur sont fictifs : ils ont été conçus à partir du tableau d'information associé au niveau Parcelle. On aboutit à la construction de deux systèmes d'échelles comparables à celui de l'étude :

- le premier, noté M-P-Z'-S', est obtenu à partir d'une classification ascendante hiérarchique et optimise l'homogénéité des niveaux d'agrégation.
- le deuxième, noté M-P-Z"-S", est obtenu à partir de partitions aléatoires et traduit *a priori* des niveaux globalement hétérogènes.

Il est délicat de fixer un critère d'homogénéité à partir des taux d'inertie suivant :

Nombre de classes	Partitions de l'étude			Partitions de la CAH			Partitions aléatoires (moyenne sur 100 simulations)		
	partit.	inertie	taux	partit.	inertie	taux	partit.	inertie	taux
925	M	2,5	100%	M	2,5	100%	M	2,5	100%
840	P	2,38	95%	P	2,38	95%	P	2,38	95%
188	Z	0,75	30%	Z'	1,6	64%	Z"	0,58	23%
16	S	0,19	8%	S'	0,7	28%	S"	0,052	2%

Tableau n°2 : *Valeur des inerties des 3 systèmes d'échelles pour un même nuage de points*

Tableau n°3 : *Décomposition de l'inertie des variables*

	totale	intra. P	intm. PZ	intm. ZS	inter. S
	2,50	0,12	1,63	0,56	0,19
MEN	0,0463	0	0,029	0,0131	0,0042
SPF	0,094	0	0,054	0,023	0,017
BAT	0,0944	0	0,0651	0,023	0,0063
MAC	0,1896	0	0,1268	0,0496	0,0132
DAC	0,1416	0	0,0973	0,0345	0,0098
OPL	0,1384	0	0,0349	0,0467	0,0568
CTB	0,0945	0	0,0651	0,022	0,0074
PRS	0,1424	0,01	0,1019	0,0269	0,0036
SEX	0,0476	0,0034	0,0345	0,0088	0,0009
AGE	0,1893	0,0152	0,1268	0,0417	0,0056
STM	0,1893	0,0165	0,1326	0,0358	0,0044
ETH	0,0466	0,0026	0,0321	0,0082	0,0037
SCL	0,141	0,009	0,0971	0,0281	0,0068
DRP	0,1423	0,0081	0,0937	0,0307	0,0098
MIG	0,0947	0,0061	0,065	0,0178	0,0058
DRO	0,1894	0,0176	0,132	0,0344	0,0054
ACT	0,1878	0,0092	0,1308	0,0403	0,0075
SMN	0,0471	0,0028	0,0297	0,0095	0,0051
LRO	0,0943	0,0067	0,0616	0,0204	0,0056
STO	0,0932	0,0091	0,0595	0,0201	0,0045
ORG	0,0949	0,006	0,0609	0,0212	0,0068

Tableau n°3a : *Valeur de l'inertie des variables*

	totale	intra. P	intm. PZ	intm. ZS	inter. S
	1000	1000	1000	1000	1000
MEN	19	0	18	23	22
SPF	38	0	33	41	91
BAT	38	0	40	41	33
MAC	76	0	78	89	69
DAC	57	0	60	62	52
OPL	56	0	21	83	299
CTB	38	0	40	39	39
PRS	57	83	63	48	19
SEX	19	28	21	16	5
AGE	76	127	78	74	29
STM	76	138	81	64	23
ETH	19	22	20	15	19
SCL	57	75	60	50	36
DRP	57	68	57	55	52
MIG	38	51	40	32	31
DRO	76	147	81	61	28
ACT	76	77	80	72	39
SMN	19	23	18	17	27
LRO	38	56	38	36	29
STO	38	76	37	36	24
ORG	38	50	37	38	36

Tableau n° 3b: *Pourcentage en colonne, CTR1*

	intra. P	intm. PZ	intm. ZS	inter. S
	5	65	22	8
MEN	0	63	28	9
SPF	0	57	24	18
BAT	0	69	24	7
MAC	0	67	26	7
DAC	0	69	24	7
OPL	0	25	34	41
CTB	0	69	23	8
PRS	7	72	19	3
SEX	7	72	18	2
AGE	8	67	22	3
STM	9	70	19	2
ETH	6	69	18	8
SCL	6	69	20	5
DRP	6	66	22	7
MIG	6	69	19	6
DRO	9	70	18	3
ACT	5	70	21	4
SMN	6	63	20	11
LRO	7	65	22	6
STO	10	64	22	5
ORG	6	64	22	7

Tableau n°3c : *Pourcentage en ligne, CTR2*

où intra.P est l'inertie intra-parcelles
intm.PZ est l'inertie du passage entre les niveaux Parcelle et Zone
intm. ZS est l'inertie du passage entre les niveaux Zone et Secteur
inter.S est l'inertie inter-Secteurs

le nombre de ménages sur la parcelle	MEN	la superficie de la parcelle	SPF
le nombre de bâtiments	BAT	le mode d'acquisition	MAC
la date d'acquisition	DAC	les opérations de lotissements	OPL
le cout de construction	CTB		

le nombre de personnes dans le ménage	PRS	le sexe	SEX
l'âge	AGE	le statut matrimonial	STM
l'éthnie	ETH	le niveau d'instruction	SCL
l'activité	ACT	le statut d'occupation	STO
l'origine du chef de ménage	ORG	la date d'installation sur la parcelle	DRP
la date d'arrivée à Ouagadougou	DRO	le lieu d'arrivée à Ouaga	LRO
le nombre de migrations intra-urbaines	MIG	le type de ménage	SMN

On peut uniquement comparer les taux entre les systèmes d'échelles d'un même nuage et dire que le système d'échelles obtenu à partir de la classification ascendante hiérarchique rend compte au mieux de niveaux homogènes en fonction du critère d'agrégation choisi.

4. 3. 2. NIVEAU D'INTERVENTION DES VARIABLES

Nous nous intéressons, ici, au niveau d'intervention de chaque variable sur le système d'échelles. Pour cela, nous allons décomposer l'inertie de chaque variable en considérant les niveaux emboités (cf §3.3.4.).

On aboutit à une série de tableaux qui donnent la valeur de l'inertie décomposée (tab. n°3a), la contribution en colonne, CTR_1 (tab. n°3b), et la contribution en ligne, CTR_2 (tab. n°3c). L'indice CTR_2 permet de mesurer la perte d'inertie obtenue dans le changement d'échelle pour chaque variable. Il identifie ainsi à quel niveau une variable intervient principalement : plus l'inertie inter-classes associée à une partition (à un niveau) est importante, plus la variable contribue à caractériser le niveau d'agrégation correspondant. Inversement, plus l'inertie intra-classes est élevée, plus la variable contribue à caractériser le niveau inférieur considéré dans le passage.

A partir du tableau n°3c, nous obtenons le graphique suivant (cf fig. n°31) qui permet **une lecture simultanée des différents niveaux d'intervention pour chaque variable :**

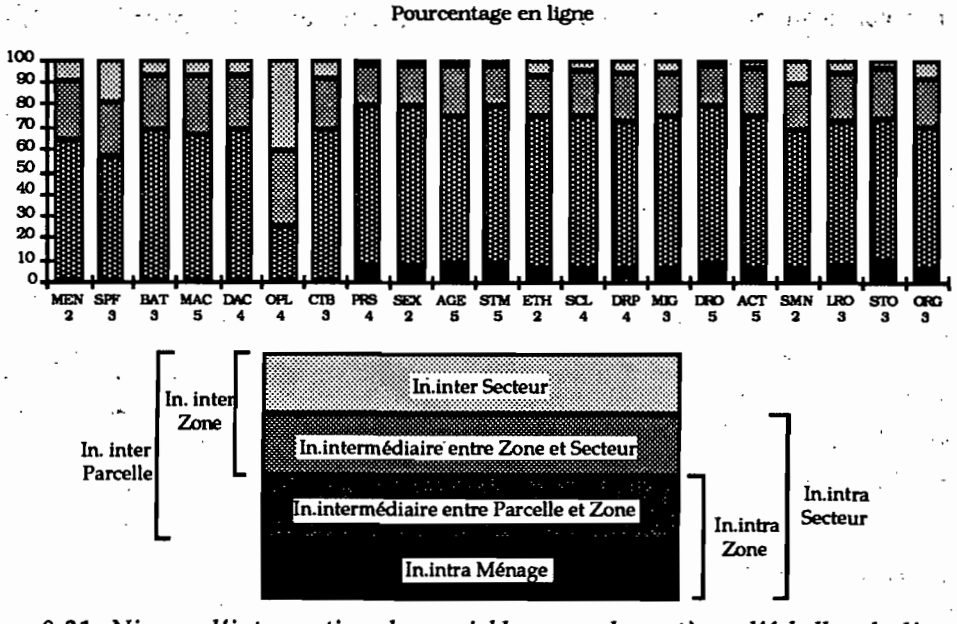


Figure n° 31: *Niveau d'intervention des variables pour le système d'échelles de l'enquête*

Les variables descriptives de la parcelle ont, en toute logique, une analyse intra-parcelles nulle (cf § 3.2.5.2). De ce fait, elles ont une inertie inter-zones ou inter-secteurs plus importante que les variables descriptives du ménage. Cependant, à l'intérieur du groupe de variables Parcelle, on observe deux catégories suivant l'importance des inerties inter-zones :

- **l'inertie inter-zones est élevée** : opérations de lotissement, superficie de la parcelle et dans une moindre mesure le groupe résidentiel. Par exemple, pour la variable "opérations de lotissement", la part d'inertie inter-secteurs est de 41% et l'inertie inter-zones est de 75% (41+34) : elle intervient essentiellement aux niveaux supérieurs. Ces variables sont directement liées aux opérations d'aménagement qui touchent souvent plusieurs parcelles d'une même zone: elles sont homogénes à l'intérieur de chaque zone de dénombrement. Cette catégorie de variables définit la morphologie de la parcelle, très liée aux opérations d'aménagement.

- **l'inertie intra-zones est élevée** : le mode et la date d'acquisition, le coût des bâtiments et leur nombre. Ces variables ont une dispersion sensiblement identique aux variables descriptives du ménage et ne sont pas en relation avec les opérations d'aménagement qui agissent aux niveaux Zone et surtout Secteur.

D'une manière générale, les variables ont une inertie intra-classes élevée pour le passage de la Parcelle à la Zone de dénombrement. Ceci confirme une certaine hétérogénéité des classes associées au découpage intermédiaire, les zones de dénombrement.

A titre de comparaison, nous reproduisons ce graphique (figure n°32) pour le système d'échelles obtenu à partir de la classification ascendante hiérarchique. L'allure de l'histogramme est très différente : les variables descriptives de la parcelle ont une variance à l'intérieur des zones et des secteurs plus faibles. Il est intéressant de remarquer que les variables "opérations de lotissement" et "superficie" qui interviennent aux niveaux Zone et Secteur (forte inertie inter-zones et inter-secteurs) ont pour ce système d'échelles une inertie inter-niveaux faible comparativement aux autres variables.

Les découpages obtenus à partir de la classification ascendante hiérarchique, c'est-à-dire du système d'échelles qui optimise l'homogénéité des niveaux supérieurs, ne tiennent pas compte de la localisation. Si l'objectif est de disposer du meilleur découpage spatial qui rend compte de niveaux supérieurs homogènes, on peut envisager de constituer le système d'échelles à partir d'une **classification ascendante hiérarchique sous contrainte de contiguité**[1].

[1][LEBART, 1978]

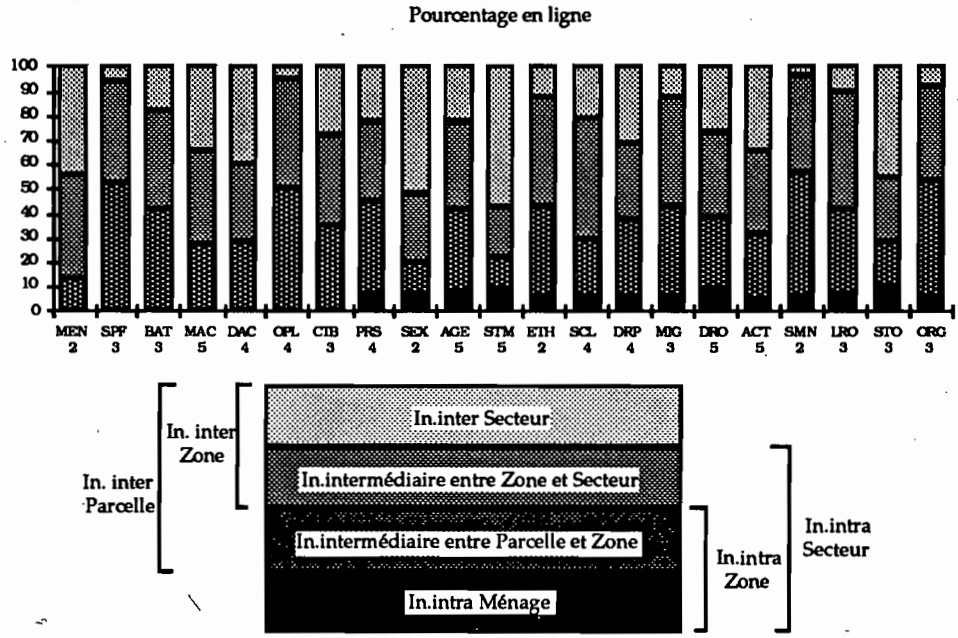


Figure n° 32: *Niveau d'intervention des variables pour le système d'échelles obtenu à partir de la Classification Ascendante Hiérarchique*

Conclusion

L'examen des plans factoriels principaux ne fait ressortir que les structures dominantes d'une organisation sociale et spatiale de la population périphérique de Ouagadougou en fonction de critères socio-économiques, d'accès au sol et au logement, d'intégration à la ville et de l'aménagement du tissu urbain. Les résultats de l'analyse du système d'échelles sont schématisés par la figure n°33. Une étude plus approfondie nécessiterait d'une part d'examiner les axes de rangs plus élevés et d'autre part de procéder par thèmes en considérant des groupes de variables mises en supplémentaires. Cette démarche, pour une interprétation plus complète, n'a pas été effectuée ici.

Cependant, l'analyse de ces plans peut être, dans un premier temps, suffisante pour dégager ce qui est déterminant dans la structuration de l'espace périphérique urbain compte tenu des variables choisies : une dispersion de la population qui ne caractérise ni un secteur particulier (à l'exception d'un type "étudiant" qui se localise en partie dans le secteur 29), ni un type d'opération systématique de lotissement. La meilleure approche de l'organisation sociale est obtenue aux niveaux Ménage et Parcelle. Au niveau Zone de dénombrement, apparaissent les types de ménages plus marginaux mais c'est surtout au niveau Secteur que sont éliminées toutes les relations entre le cycle de vie et le mode d'acquisition de la parcelle (information donnée par l'analyse du passage entre les niveaux Zone et Secteur). La progression de la participation des variables qui qualifient les niveaux est nette : les variables qui caractérisent bien les niveaux inférieurs ne sont plus des éléments de discrimination aux niveaux supérieurs et inversement.

L'analyse du système d'échelles permet alors de décider du niveau d'analyse pertinent pour un phénomène donné. Ainsi :

- **Pour comprendre l'organisation sociale et résidentielle** de la périphérie de Ouagadougou, les **niveaux Parcelle et Ménage** sont les plus indiqués et l'analyse mettra en valeur les modes d'accès au sol et au logement, l'intégration urbaine et la structure socio-économique de la population. Les principaux éléments de cette organisation sociale sont donnés par les différentes étapes du **cycle de vie.**

- **Pour étudier la politique d'aménagement urbain et ses conséquences, le niveau Secteur** est le plus adéquat et l'analyse portera sur l'évolution du tissu urbain, les caractéristiques morphologiques de l'habitat et dans une moindre mesure sur les capacités des secteurs à accueillir et retenir leur population. Globalement, on peut poser que **la localisation des principaux types de ménages de la périphérie ouagalaise est peu marquée par les opérations**

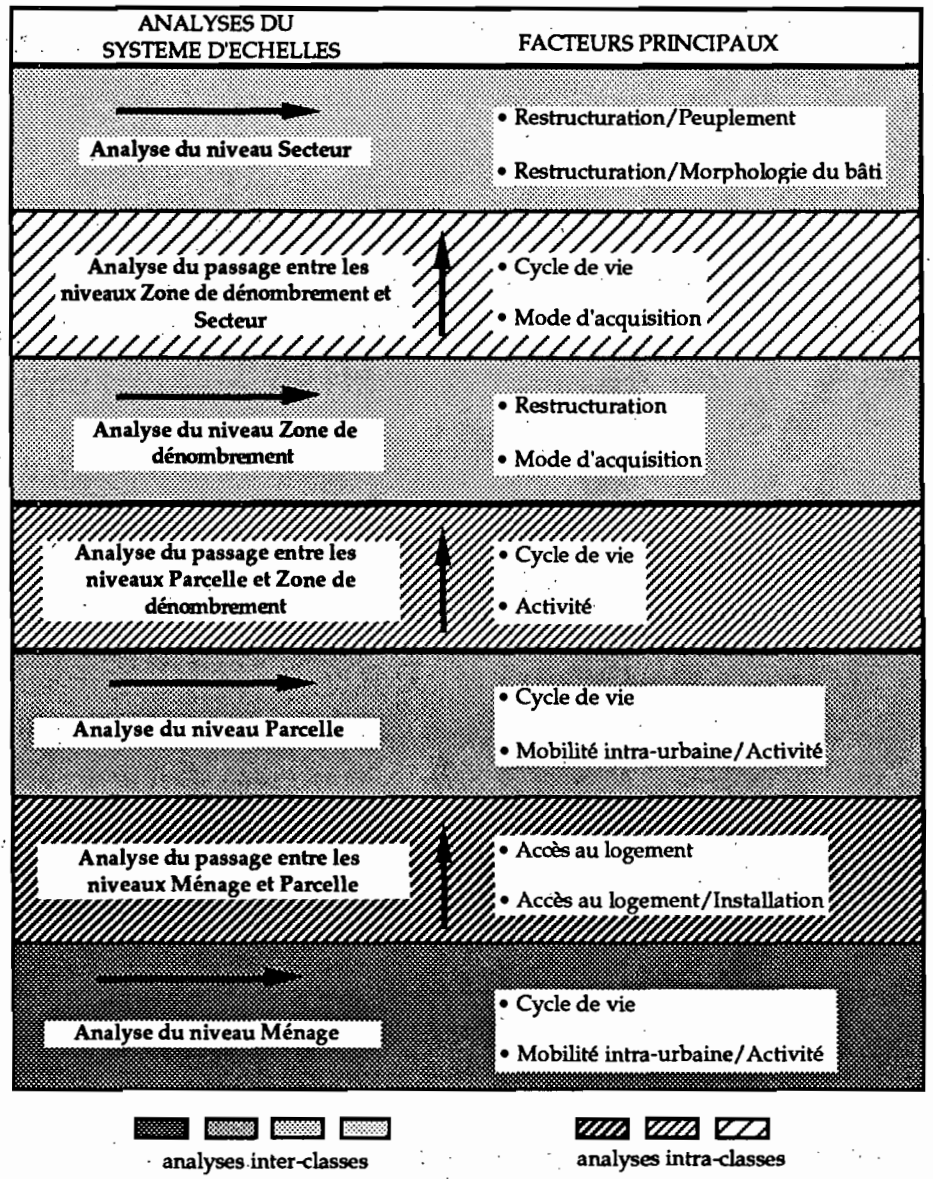


Figure n° 33 : *Synthèse de l'analyse du système d'échelles à partir de plans factoriels principaux*

étatiques d'aménagement du tissu urbain. Celles-ci ont surtout une influence sur la composition morphologique de l'habitat. On identifie cependant quelques éléments sociologiques, niveau d'instruction élevé ou installation ancienne à Ouagadougou, qui interviennent dans la structuration de l'espace à un niveau géographique plus général.

- Enfin, pour saisir **la relation entre les variables descriptives du milieu** (habitat et tissu urbain) **et celle du groupe résidentiel,** le niveau **Zone de dénombrement** est le plus adapté. Ce niveau d'analyse semble être intéressant car la structure du nuage n'est pas énormément modifiée par rapport au niveau le plus fin et si ce sont les opérations de lotissement qui déterminent surtout ce niveau puis le mode d'acquisition de la parcelle, les structures des ménages restent présentes à cette échelle, même si elles sont parfois masquées. On identifie alors des modes résidentiels "traditionnel", "moderne" et les modes "non précisés", liés

respectivement à des comportements de ménages qualifiés "d'ancien", de "moderne" et "d'étudiant".

Les opérations d'aménagement urbain se traduisent par une homogénéisation du bâti urbain mais, à la date de l'enquête (1987), elles ne provoquent ni de regroupements spatiaux sociologiques apparents, ni de stratégies perceptibles directement.

Conclusion générale

LA METHODE PROPOSEE

L'analyse statistique d'un système d'échelles que nous avons développée est basée sur les techniques descriptives d'analyses multivariées des variables qualitatives. Elle consiste dans un premier temps à exploiter simultanément des informations définies sur des niveaux d'observation différents et dans un deuxième temps à observer d'une part chaque niveau d'analyse, d'autre part les changements de niveaux. Deux problèmes méthodologiques se posent pour :

- **la mise en relation d'informations issues de niveaux différents :**
  la conception du tableau d'information doit d'une part répondre à une structure rectangulaire de données homogènes et d'autre part respecter la structure hiérarchique de l'information. Deux choix sont à faire : l'un sur la population statistique et l'autre sur les niveaux d'analyse. En effet, l'existence de plusieurs types d'unités d'observation entraîne l'existence de plusieurs populations statistiques potentielles et induisent diverses unités d'analyse.
  Dans l'optique de généraliser la construction d'un tableau d'information adapté à toutes les situations possibles suivant les choix à faire, nous avons mis en évidence un **système de pondération** cohérent fonction d'une population statistique de référence qui aboutit à une série de tableaux, chacun associé à un niveau d'analyse. Nous pouvons ainsi, par **des relations d'agrégation ou d'affectation** qui déterminent la relation entre le niveau de définition des variables et le niveau d'analyse, imaginer tous les croisements possibles entre les groupes de variables.

- **l'analyse du système d'échelles** proprement dit :
  il s'agit de dégager à la fois les structures spécifiques de chaque niveau et celles que l'on perd dans le changement d'échelle.
  Nous proposons une méthode **basée sur l'analyse des correspondances multiples et sur la décomposition de l'inertie** qui consiste en une **succession d'analyses des niveaux d'agrégation** (analyses inter-classes) **et d'analyses des passages entre deux niveaux** (analyses intra-classes) ; celles-ci jouent le rôle d'aides à l'interprétation du système d'échelles et permettent de répondre à certaines questions posées par le processus d'agrégation.

Cette méthode d'analyse d'un système d'échelles est simple à mettre en œuvre puisque tout repose sur la construction de tableaux sur lesquels il s'agit d'appliquer une analyse factorielle des correspondances classique.

Confronter des variables définies sur des niveaux différents et travailler sur plusieurs échelles offrent des possibilités pour de nouveaux traitements de l'information, tels que la prise en compte d'unités collectives en démographie, ou la mise en relation de données descriptives du ménage et de celles relatives aux établissement en économie.

Si on intègre une composante géographique, cette analyse statistique constitue alors une méthode performante d'**analyse spatiale exploratoire,** qui trouve un champ d'application privilégié dans le traitement de données intégrées dans un système d'information géographique. Maîtrise des processus d'agrégation et/ou de ventilation spatiales, mise en évidence des structures spécifiques aux différents niveaux d'information et des relations entre ces niveaux, réflexions sur la pertinence et l'effet des découpages géographiques, constituent les principales applications de cette méthode qui permet d'aller plus avant dans l'exploitation combinée d'informations correspondant à des découpages spatiaux différents.

LES LIMITES ET LES PERSPECTIVES D'AMELIORATION

L'analyse du système d'échelles reste cependant à développer dans trois domaines d'application :

- **pour le traitement de plus de deux groupes de variables définies sur des niveaux différents.**
  Nous venons de mettre en place des tableaux d'information qui permettent de croiser des variables définies sur des niveaux différents. Cependant si, dans une analyse des correspondances multiples, les groupes de variables (de nature différente) semblent intervenir à part égale,*"cette égalité apparente peut cacher un déséquilibre important"*[1]. Pour le croisement de deux groupes de variables, l'analyse d'un sous-tableau de Burt permet de faire jouer un rôle symétrique aux deux groupes, que ce soit à un niveau donné ou dans le passage d'un niveau à un autre. Il est ainsi possible de regarder les interactions qui existent entre ces groupes, ceci à n'importe quelle étape du changement d'échelle. Mais lorsque l'on dispose au moins de trois groupes de variables, on peut se poser la question de savoir si l'analyse des correspondances multiples permet toujours d'appréhender de manière symétrique l'ensemble des variables de nature différente. Sachant que dans le cas de variables issues d'objets divers, d'emblée l'inertie des variables est influencée par son niveau de définition, on peut s'interroger sur la manière de mesurer la part d'inertie non significative inhérente à chaque niveau d'analyse.
  Dans cette optique, l'analyse factorielle multiple, réalisée sur les différents types de tableaux d'information mis en évidence dans l'analyse du système d'échelles, peut être une orientation à prendre. En effet, d'une manière générale, cette méthode *"est conçue spécialement pour l'analyse de tableaux comportant plusieurs groupes de variables. Elle transgresse ces limites et permet surtout une comparaison systématique des groupes à travers*

[1][ESCOFIER ET PAGES, 1988] p104 .

l'ensemble de points de vue très divers que la richesse de la structure de groupes de variables suggère"[1].

- **pour le traitement de variables quantitatives.**
  La méthode proposée est adaptée aux variables qualitatives ou considérées comme telles et repose de ce fait sur l'analyse factorielle des correspondances qui utilise la distance entre profils, celle du khi-deux. Mais, il est possible d'envisager une méthode similaire d'analyse d'un système d'échelles appropriée à des variables quantitatives.
  On peut envisager, sur le même principe, une méthode basée sur l'analyse en composantes principales (laquelle utilise la distance euclidienne) ce qui ne change en rien le principe de décomposition de l'inertie puisqu'il ne dépend pas de la nature de la distance. Toutefois, pour concevoir l'analyse intra-classes, les variables doivent être centrées et réduites non pas par rapport à la moyenne et l'écart-type effectués sur la population totale, ce qui est le cas pour une analyse en composante principale classique, mais sur la population définie par chaque classe. Il s'agit alors d'une part de définir un critère d'homogénéité qui ne peut plus être basé sur la notion de profils semblables et d'autre part d'étudier les conséquences pour l'interprétation. En effet, on ne peut plus utiliser le principe de l'équivalence distributionnelle pour réaliser cette analyse basée sur l'analyse en composantes principales.

- **pour un indicateur d'homogénéité.**
  La méthode d'analyse d'un système d'échelles reste limitée à son univers d'étude. L'étape suivante serait alors de réfléchir à la conception d'un indice qui permettrait la comparaison entre différents systèmes d'échelles. Pour cela on peut s'interroger sur la construction d'un indicateur d'homogénéité des niveaux qui permettrait, de plus, d'étalonner le niveau d'analyse pertinent pour stratifier ou pour cartographier.
  Concevoir cet indicateur à partir des taux d'inertie peut être une première piste mais il faut auparavant établir des tests statistiques pour connaitre la part d'inertie due au découpage et définir s'il est possible de comparer par cet indicateur des découpages entre eux. En effet, dans le cadre d'analyses multivariées, ces taux d'inertie dépendent du nombre de variables et donc du découpage.

[1][ESCOFIER ET PAGES, 1988] p109.

Annexe 1

TECHNIQUES D'ENQUETE

Annexe 1.1 : L'échantillon

Constitution de la base de sondage

La constitution d'une base de sondage reste souvent problématique dans les villes à fort taux d'accroissement où l'habitat spontané, spécifique des périphéries, se développe sans contrôle et risque, dans le cas de Ouagadougou, d'être bouleversé, ce qui rend difficile une lecture instantanée de la ville au moment de l'enquête.

Au regard de la documentation, deux possibilités s'offraient pour l'élaboration de la base de sondage que J. DESABIE[1] définit comme : une liste exhaustive, récente, sans répétition et permet d'identifier facilement l'unité statistique choisie c'est-à-dire, dans notre cas, la parcelle.

A partir des photographies aériennes

Au moment de la conception de l'enquête, en 1984, une première initiative était de constituer cette base à partir d'une mosaïque de photographies aériennes à l'échelle de 1/10000 datant de 1982, afin d'effectuer un sondage aréolaire. Pour ce faire, la périphérie a été quadrillée par carrés d'un hectare. Seuls les carrés comportant au moins une parcelle ont été retenus, soit 5241 éléments, constituant ainsi la base de sondage. Les unités aréolaires ont été sélectionnées suivant la méthode du tirage systématique et chaque parcelle appartenant (dont plus de la moitié pour les parcelles "à cheval".) à ces unités compose l'échantillon.

Il est certain que ces unités étaient de densités très inégales quant au nombre de parcelles, puisque le tissu urbain est à cet égard très différencié. Il s'agissait alors d'établir des zones d'habitat homogènes, c'est-à-dire de définir des strates afin d'avoir une bonne répartition de l'échantillon sur chacune d'entre elles, sans privilégier une zone d'habitat. Quatre types de zones d'habitat ont été retenus et cartographiés:

- zone d'habitat lotie,
- zone d'habitat dense non lotie.
- zone d'habitat peu dense non lotie,
- zone d'habitat discontinue non lotie.

Cependant le principal défaut de cette méthode, dans notre cas, a résulté de l'ancienneté des documents disponibles qui rendait le travail très délicat en particulier dans les quartiers à habitat mal structuré, la majeure partie des constructions actuelles n'y figurant pas. Il a donc fallu noter par la suite toutes les particularités permettant de retrouver aisément les lieux pour évaluer les modifications (d'autant plus importantes que la politique récente d'urbanisation a profondément transformé la périphérie) et l'extension des constructions spontanées qu'elle a connues en quatre années. Par conséquent réactualiser les photographies aériennes engendrait un certain nombre d'erreurs non redressables affectant la représentativité de l'échantillon.

Ceci nous a donc amené à adaopter une autre technique pour établir la base de sondage.

A partir du recensement général de la population de 1985

L'enquête, qui avait été initialement prévue avant le recensement général de la population (RGP) effectué en décembre 1985 par l'Institut National de la Statistique et de la Démographie (INSD), a été, pour diverses raisons administratives, reportée en début d'année 1986. Ainsi espérions-nous disposer d'une base de sondage répondant aux critères demandés.

Les modalités de la préparation de l'échantillon axée, par conséquent, sur un sondage aréolaire ont

---

[1][DESABIE,1966]

été modifiées par l'existence d'une base de sondage presque satisfaisante obtenue à partir du dernier recensement général de la population. En effet, les agents recenseurs de l'INSD ont procédé, par zone de dénombrement (plus petite unité spatiale dont l'ensemble forme une partition du secteur et *a fortiori* de l'espace périphérique) et suivant un itinéraire imposé, à un numérotage continu des parcelles susceptibles d'être habitées. Une première approche sur le terrain nous a permis de relever, auprès des délégués CDR de chaque secteur, le nombre de parcelles numérotées (qui ne correspond pas toujours au nombre de parcelles occupées) et, dans certains secteurs, le nombre de parcelles enquêtées (c'est-à-dire habitées) au moment du recensement. Cependant, ne pouvant, dans un premier temps, considérer que les parcelles numérotées, la base de sondage reste imparfaite. Nous verrons que l'échantillon a du être réajusté pour ne plus prendre en compte que les parcelles habitées au moment de l'enquête.

En raison, d'une part, des stratégies foncières, (telles que la construction de bâtiments non occupés sur des parcelles de la périphérie en vue d'une appropriation du terrain créée par un droit d'usage), et, d'autre part, de la politique urbaine entraînant des déguerpissements (effectués pour certains entre le recensement et l'enquête), nous avons été trouvés confrontés à une sur-représentation des parcelles habitées ou susceptibles de l'être. De nombreux retours sur le terrain ont été nécessaires lors de la constitution de l'échantillon pour mesurer ces phénomènes. Ainsi par exemple, en quelques jours, une zone du secteur 30 a été partiellement déguerpie pour la constitution d'une zone administrative, affectant environ 200 parcelles. Il a fallu tenir compte de la densité réelle de l'habitat sur l'espace périphérique au moment de l'enquête et redresser éventuellement l'échantillon élaboré par rapport à une base de sondage constituée trois mois auparavant.

Il est à souligner que le fait de disposer d'une information par zone de dénombrement facilite ce redressement et permet ainsi de réduire les éventuelles erreurs. Rappelons que la zone de dénombrement est une notion spatiale totalement arbitraire, considérée par l'INSD comme une aire géographique dont le seul objectif est de faciliter le déroulement du recensement en regroupant théoriquement 800 personnes en moyenne.

Avantages et inconvénients de ces deux techniques d'élaboration d'une base de sondage

Au niveau d'une logique urbaine, la base de sondage cartographique repose sur des types d'habitat, caractéristiques non négligeables pour une étude urbaine sur Ouagadougou, qui permettent une meilleure définition de l'espace que ne le propose l'unité spatiale et administrative issue du RGP: le secteur (et la zone de dénombrement). En effet, se référer à des types de tissus urbains correspondant aux différentes étapes du lotissement systématique des quartiers spontanés de la périphérie ouagalaise aurait été préférable et plus satisfaisant pour un urbaniste plutôt que de se baser sur des zones de recensement essentiellement construites sur la capacité moyenne de travail d'un agent recenseur. Cependant, nous avons élargi cette zone au secteur qui reste malgré tout un découpage administratif, mais conserve un sens dans l'étude de la politique urbaine de Ouagadougou puisque ces secteurs sont une résultante de cette politique et que la gestion des secteurs est un des aspects de cette étude.

Mais, nous avons vu qu'il existait une discordance entre les photographies aériennes disponibles et la réalité du terrain entraînant des risques d'erreurs importants, accentués par une identification difficile des parcelles. C'est pourquoi, la base de sondage obtenue à partir du RGP semble préférable pour la qualité des résultats et est en mesure d'assurer une bonne représentativité de l'échantillon, aspect fondamental sur un plan purement statistique.

Dans les deux cas, il n'existe pas d'informations supplémentaires portant sur tous les éléments de la base de sondage, permettant, s'il y a lieu, d'améliorer la précision des estimations, d'autant moins que la plupart d'entre elles se réfèrent à des sous-populations. En effet, il ne nous a pas été possible d'avoir accès aux données du RGP, en rapport avec notre échantillon, du fait d'un écart de temps trop réduit entre l'exécution du recensement et le lancement de notre enquête et, à partir de la base de sondage cartographique, nous ne pouvons disposer d'emblée de telles données.

Mise en place de l'échantillon

Taille de l'échantillon

Nous ignorons le degré d'homogénéité de la population et par conséquent des sous-populations qui vont nous intéresser dans cette étude (sous-populations issues d'une éventuelle stratification, des statuts d'occupation, des types de mobilités de ménages, des individus composants le ménage). Aussi il est préférable de travailler sur un échantillon important afin de donner une bonne précision des estimations pouvant porter sur ces sous-populations (cette précision dépendant de la taille de l'échantillon et non de la taille de la population).

Par ailleurs, nous avons fixé l'effectif non pas par rapport à la parcelle mais par rapport au ménage soit un effectif de 1000 ménages. En effet, en pratique, optimiser la taille de l'échantillon dépend non seulement des objectifs scientifiques mais aussi des crédits disponibles. Or le coût de la phase de l'enquête sur le terrain dépend :

- du coût du questionnaire (matériel),
- du temps passé pour repérer la parcelle,
- du temps passé pour l' interview (éventuellement plusieurs visites si le chef de ménage est absent).

Par rapport à ces contraintes, le coût final a été calculé sur la base du questionnaire relatif au chef de ménage et non pas sur la base de la parcelle.

Afin d'approcher les 1000 ménages et en supposant une moyenne de 1,2 ménages par parcelle, 850 parcelles est l'effectif retenu pour l'échantillon. (Cette moyenne est approximative et a été réajustée puisque basée sur les premiers résultats provisoires du RGP. Elle dépend, de plus, de la définition du ménage donné par l'INSD qui est différente de la nôtre).

Finalement 853 parcelles ont été enquêtées sur l'ensemble de la périphérie impliquant l'interview de 938 ménages lors du premier passage. Mais ce résultat n'apparaissant qu'au terme du premier passage de l'enquête, il ne nous était plus possible de réenquêter d'autres parcelles, la composition de l'échantillon ne pouvant plus être modifiée d'une manière homogène.

Taux de sondage

Bien que nous puissions apprécier, pour chaque secteur, le nombre de parcelles enquêtées et donc occupées au moment de recensement, la base de sondage prend en compte, dans sa numérotation, l'ensemble des parcelles habitées ou non, soit 40091 parcelles sur la globalité de la périphérie. Or l'échantillon doit être tiré à partir d'une liste exhaustive constituée de parcelles occupées au moment de l'enquête. Les interventions sur le terrain nous ont permis d'évaluer les parcelles détruites depuis le recensement (prise de connaissance de vastes zones ayant données lieu à des déguerpissements), et d'estimer alors à 36888 le nombre de parcelles occupées ou susceptibles de l'être. Cependant, effectuer le tirage de l'échantillon à partir du nombre de parcelles habitées aurait déqualifié la base de sondage, excluant par la même les dernières parcelles numérotées et habitées (nous n'étions pas en mesure d'identifier, à partir de cette liste, les parcelles non occupées).

Se pose, alors, le problème du tirage d'une parcelle ayant disparu au moment de l'enquête ou inoccupée et non identifiable à priori d'après la liste puisque nous sommes dépendants de la numérotation, seul moyen permettant de repérer les individus. Deux possibilités s'offrent pour construire l'échantillon à partir de cette base:

- soit maintenir l'effectif de l'échantillon et si une parcelle sélectionnée est inexistante ou inoccupée, enquêter la parcelle habitée la plus proche. Celle-ci n'a donc pas la même probabilité d'être choisie: une parcelle se situant à proximité de parcelles ne répondant pas aux critères demandés a plus de chances d'être sélectionnée ;

- soit augmenter l'effectif de l'échantillon pour tenir compte des parcelles inhabitées qui risquent d'être tirées. De ce fait, les zones de dénombrement dans lesquelles se trouvent peu de parcelles inhabitées sont sur-représentées (moins de chances de sélectionner ce type de parcelles). De même, les zones où le nombre de ces parcelles est plus important, seront sous-représentées (plus de chances de tirer une parcelle inhabitée).

Nous avons opté pour cette deuxième solution. En effet, il est ainsi possible d'effectuer un tirage équiprobable, à partir de la base de sondage fournie par le RGP, la probabilité de choisir une parcelle habitée ou non étant connue et non nulle. Il s'agira alors de redresser l'échantillon en fonction d'un état des lieux au moment de l'enquête en retranchant (ou ajoutant) le nombre voulu d'unités dans les zones sur (ou sous) représentées par le tirage au sort. Comme il est plus simple de soustraire que d'ajouter, il est donc souhaitable de posséder, au départ, un échantillon plus important et de le pondérer, *a postériori*, pour respecter sa répartition[1].

Nous avons donc fixé à 1000 le nombre de parcelles devant être sélectionnées, soit un taux de sondage de 1/40 par rapport à l'ensemble des parcelles numérotées lors du recensement.

Plan de sondage

Stratification

La base de sondage a été décomposée suivant les 16 secteurs administratifs retenus pour notre étude. Nous avions deux procédés possibles pour effectuer le tirage:

- soit sur la base de sondage considérée comme la réunion des zones administratives,
- soit sur les bases de sondage obtenues pour chaque secteur, indépendamment les uns des autres.

Le dernier cas relève d'un tirage à deux degrés au moins, les unités primaires étant les circonscriptions administratives. Il s'agit d'un tirage stratifié, les strate correspondant aux secteurs. En l'occurrence, nous avons retenu tous les secteurs périphériques puisque définissant le terrain d'étude.

La stratification était donc implicite ce qui a permis d'assurer une bonne représentativité de l'échantillon et améliore généralement la précision des estimations. Par ailleurs, le fait de stratifier suivant la variable spatiale, le secteur, permettait un contrôle de l'espace périphérique: une erreur d'observation sur une zone de dénombrement, et *a fortiori* sur un secteur, n'entraîne pas la même erreur sur les autres secteurs.

Cette précaution était d'autant plus importante que pour chaque secteur, nous ignorions le nombre réel de parcelles habitées. Il aurait été difficile de réajuster l'échantillon si tous les secteurs avaient été confondus.

De plus, l'extension urbaine répond à une logique géographique. L'identification du lieu de l'extension est un élément fondamental d'analyse de la ville. La situation spatiale d'une parcelle est une information sûre et stable qui reste en liaison forte avec le phénomène étudié, même si elle ne répond pas à un concept urbain. Nous avions, par conséquent tout intérêt à exploiter le secteur comme critère de stratification.

En adoptant un taux de sondage constant pour chaque secteur,1/40, la notion de densité de l'habitat est respectée à l'intérieur de chaque secteur et sur l'ensemble de la périphérie et tout l'espace étudié est couvert (proportionnalité de l'effectif des parcelles entre l'échantillon et la population : il s'agit alors d'une allocation proportionnelle). De plus, "*un plan stratifié où les taux de sondage sont les mêmes* (allocation proportionnelle) *dans chaque strate ne charge que des échantillons*

[1][DESABIE,1966]

représentatifs"[1]. Par conséquent, "*un échantillon stratifié représentatif possède donc, et possède seul, la propriété de pouvoir être dépouillé comme un recensement*"[2]. "*Cela vaut rarement la peine d'utiliser un échantillon de Neyman* (allocation optimale) *lorsque l'on cherche à estimer des proportions*"[2], comme cela est notre cas.

Nous assurons dès maintenant la constitution d'un échantillon représentatif mais il importe alors de décider du mode de tirage proprement dit à l'intérieur de chaque strate tout en respectant une bonne répartition des parcelles sélectionnées.

Tirage systématique

Soulignons que l'INSD, lors de la préparation du recensement, a attribué à chaque parcelle un identifiant. Ce dernier rend compte de la localisation géographique de la parcelle puisqu'il est fonction du secteur et de la zone de dénombrement permettant ainsi un repérage précis des unités à enquêter. L'échantillon a été tiré par secteur sur l'identifiant affecté à chaque unité statistique. Nous avons envisagé un tirage indépendant dans chaque strate.

Le choix d'un tirage systématique, qui consiste à partir d'une parcelle prise au hasard à tirer les autres parcelles suivant un intervalle défini par le taux de sondage, est dans notre cas souhaitable puisque les parcelles qui occupent les rangs voisins dans la numérotation ont tendance à se ressembler par rapport à leur situation de lotissement ou à un type d'habitat. Or un des objectifs dominants est l'appréciation de l'impact des opérations de lotissement sur les caractéristiques de la population. De plus, un tirage systématique peut être employé sans ambiguïté dans notre cas, aucune régularité quelconque ne pouvant coïncider avec le pas choisi pour le sondage.

Ce type de tirage assure une répartition géographique satisfaisante de l'échantillon par secteur et ne favorise pas une situation résidentielle plutôt qu'une autre.

Ceci tient à l'approximation du tirage systématique à un tirage équiprobable sans remise. Si l'indice de l'individu n'apporte pas d'information par rapport au caractère étudié, tirer de n'importe quelle façon un échantillon équivaut intuitivement à le tirer de façon équiprobable sans remise. Cette technique procède, donc , par balayage homogène du registre des unités qui s'y prêtent (puisque ne présentent pas de périodicités par rapport aux caractères étudiés).

Le tirage systématique fournit par conséquent une méthode simple à mettre en oeuvre pour approcher le tirage équiprobable sans remise et ne privilégie aucune zone urbaine[3].

Le même mode de tirage a été appliqué à chaque secteur, où les valeurs d'initialisation ont été choisies aléatoirement, indépendamment les unes des autres, pour respecter l'indépendance des tirages entre les strates.

Conclusion

Le choix du plan de sondage proposé a été restreint par les nombreuses contraintes rencontrées mais reste, malgré tout, classique en assurant une bonne représentativité.

Il s'agit donc d'un plan stratifié qui peut être envisagé sous deux aspects suivant la considération de l'unité statistique d'enquête:

- soit comme un sondage à deux degrés si l'on considère la parcelle comme individu statistique. Les unités primaires sont les strates identifiées aux secteurs et le tirage au premier degré est un recensement (tous les secteurs périphériques sont pris en compte). Au second degré, les

[1][GOURIEROUX,1981]
[2][DESABIE,1966] p145 et 147.
[3][GOURIEROUX,1981]

parcelles sont tirées selon un plan équiprobable sans remise approché par un tirage systématique avec un taux de sondage approximativement constant (les éventuelles variations proviennent du choix aléatoire des premières parcelles prises au hasard pour chaque secteur);

- soit comme un plan à trois degrés si l'on considère le ménage comme individu statistique. Les unités primaires sont les secteurs, les unités secondaires, les parcelles et les unités tertiaires, les ménages résidant sur les parcelle sélectionnées. Les tirages au premier et troisième degrés étant des recensements, le tirage au deuxième degré est un tirage équiprobable sans remise. Il s'agit alors d'un sondage stratifié en grappes où une parcelle définit une grappe de ménages.

Redressement de l'échantillon

La technique d'échantillonnage retenue a été appliquée en deux temps. Chaque strate, et *a fortiori* l'échantillon global, a été initialement sur-représentée. Puis, suivant un état des lieux effectué simultanément à l'enquête pour évaluer approximativement le nombre de parcelles déguerpies, des unités statistiques ont été soit éliminées d'elles-même car détruites ou inhabitées, soit rejetées par un tirage au sort si le nombre d'unités tirées était trop important.

Nous disposons, comme uniques données, du nombre de parcelles numérotées par secteur et d'un nombre approximatif de parcelles recensées et donc habitées donnant par conséquent les effectifs d'échantillon suivants (cf tableau n°4):

- l'effectif E1 sur-représenté issu de la base de sondage de parcelles numérotées avec un taux de 1/40 d'où est tiré l'échantillon initial soit 1000 parcelles ;
- l'effectif E2 théorique issu du nombre de parcelles habitées (recensées) ou susceptibles de l'être, obtenu d'après un état des lieux effectué lors de la préparation de l'échantillon, avec un taux de sondage de 1/43 afin d'obtenir 850 parcelles ;
- l'effectif E3 observé correspondant au nombre de parcelles enquêtées soit 853 parcelles constituant l'échantillon réel.

Secteurs	Base de sondage	Echantillon sur-représenté	Nombre de parcelles		Echantillon théorique	Echantillon observé		coef.de pondér.
	parc. numérot	(E1) 1/40	recensées	%	(E2) 1/43	(E3)	%	E2 /E3
S15	3497	87	3325	9	77	73	8,6	1,05
S16	4469	111	4158	11,3	96	96	11,2	1,00
S17	5801	145	4936	13,4	114	111	13	1,03
S18	735	18	614	1,7	14	18	2,1	0,78
S19	2094	52	2023	5,5	47	45	5,3	1,04
S20	946	24	881	2,4	20	21	2,5	0,95
S21	917	23	782	2,1	18	21	2,5	0,86
S22	2851	71	2664	7,2	61	59	6,9	1,03
S23	4291	107	4023	10,9	93	89	10,4	1,04
S24	1573	39	1415	3,8	32	32	3,7	1,00
S25	917	23	890	2,4	20	22	2,6	0,91
S26	299	7	299	0,8	7	7	0,8	1,00
S27	1744	44	1684	4,6	39	41	4,8	0,95
S28	3368	84	3126	8,5	72	68	8	1,06
S29	3632	91	3475	9,4	80	76	8,9	1,05
S30	2957	74	2593	7	60	74	8,7	0,81
Ens.	40091	1000	36888	100	850	853	100	

Tableau n°4 : *Redressement de l'échantillon*

Un redressement de l'échantillon a été opéré *a postériori* par l'affectation d'un coefficient de pondération (défini suivant le rapport E2/E3) à chaque parcelle, en fonction de son appartenance au secteur. La moyenne, dans l'échantillon, du coefficient de pondération obtenu est proche de l'unité puisque l'effectif observé n'est pas égal à l'effectif théorique; leur différence est de trois parcelles (qui ont été enquêtées et qu'il n'était pas nécessaire d'éliminer). Les valeurs extrêmes sont 0,78 et 1,06 sachant que 77% des individus ont un poids supérieur ou égal à 1.

Cette pondération permet de se rapprocher de l'effectif théorique dont nous ignorons réellement la valeur. Nous pouvons, par conséquent, nous demander dans quelle mesure ce redressement a une portée significative sur les résultats, d'autant plus qu'il s'agit d'une correction assez légère ; ce redressement a pour objectif de corriger les inégalités dans les probabilités de sélection des différentes parcelles et d'assurer par la même, une bonne répartition géographique des unités statistiques. Or nous nous apercevons que l'ordre de grandeur des pourcentages de cette répartition n'est pas altéré. Seul le secteur 30 a été sur-estimé.

Nous avons, cependant, maintenu le coefficient de pondération puisqu'il permet malgré tout une meilleure approximation de la réalité. Il aurait cependant été intéressant de mesurer les conséquences d'une telle pondération sur les estimations. Pour cela il aurait fallu disposer des mêmes données en provenance d'une autre source suivant des critères identiques à notre étude.

Conclusion

La figure n°34 récapitule les différentes étapes de la mise en œuvre de l'échantillon. Notre souci étant de posséder de bonnes précisions, l'échantillon tel qu'il a été conçu reste représentatif. Il s'agit seulement de respecter un effectif suffisamment élevé lors de l'étude de sous-populations. De plus, il semble inutile de vouloir donner une grande précision aux pourcentages au delà de leurs valeurs numériques entières.

Il ne s'agit pas réellement d'un sondage probabiliste, bien qu'il en soit rapproché puisque la base de sondage, telle qu'elle fut élaborée, ne pouvait donner aux parcelles la même probabilité d'être sélectionnées : soit nous avons éliminé des parcelles de la liste si le secteur avait subi peu de bouleversements, soit elles s'éliminaient d'elles-mêmes. Nous n'étions pas en mesure d'estimer le nombre exact de parcelles habitées par secteur. De ce fait, on ne peut connaître la précision et le risque d'erreur des paramètres calculés sur l'échantillon et étendus à la population-mère.

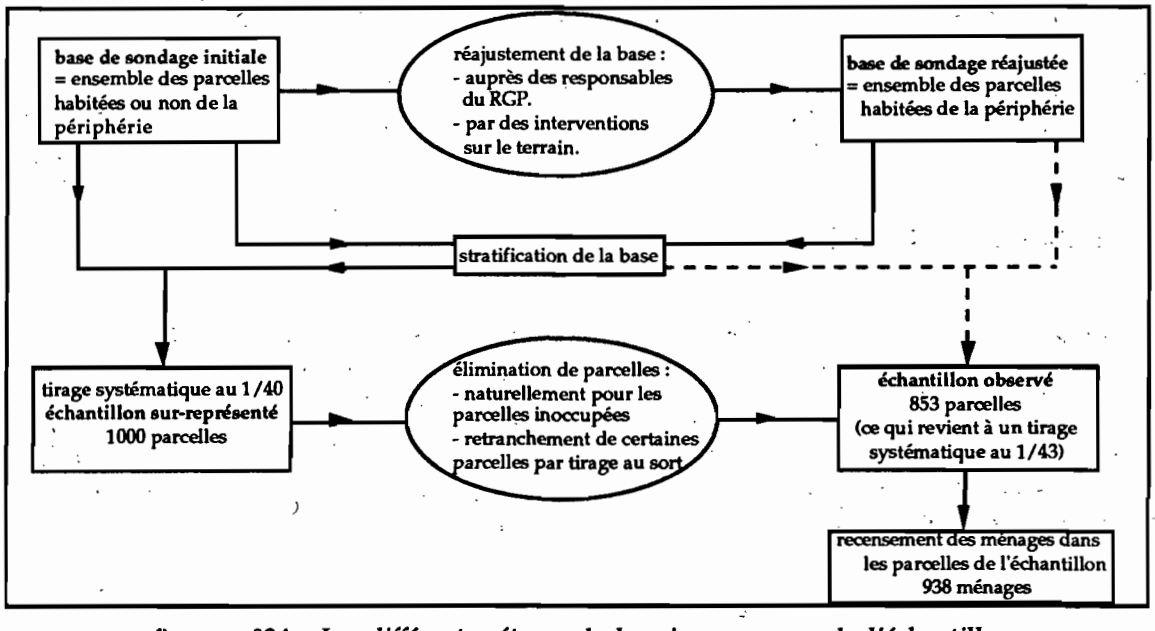


figure n°34 : *Les différentes étapes de la mise en œuvre de l'échantillon*

Annexe 1.2 : Les questionnaires

Le premier passage

Le questionnaire correspondant se décompose en trois parties:

1- détermination des caractéristiques du ménage: caractéristiques socio-économiques du chef de ménage et des individus composant le ménage (épouses résidentes et non résidentes, enfants, dépendants), migrations du chef de ménage, et équipements du ménage (p 1 à 11-16);

2- détermination des modalités d'occupation et d'installation du chef de ménage sur la parcelle (p 12-14-15);

3- détermination des caractéristiques de la parcelle (ses modalités d'acquisition, son équipement), et des caractéristiques des bâtiments (p 13-16-17);

Le principe du questionnaire est de comporter deux parties fixes (1 et 2) et une partie mobile (3) uniquement destinée au chef de ménage responsable de la parcelle.

Le deuxième passage

Le deuxième passage a nécessité trois questionnaires liés aux trois situations possibles:

- le ménage présent lors du premier passage a quitté la parcelle. Un questionnaire plus restreint concerne surtout la parcelle mais tente de recueillir quelques informations sur ce ménage.

- le ménage a déjà été enquêté lors du premier passage. Ce questionnaire concerne principalement la mobilité interne au ménage et les éventuels changements économiques, sociaux et fonciers survenus dans l'année.

- le ménage est arrivé sur la parcelle depuis le premier passage. Le questionnaire a repris, en l'allégeant, la structure de celui du premier passage, excluant les migrations du chef de ménage et les caractéristiques des épouses et des enfants. Il contient une partie rétrospective légère correspondant à sa situation précédant son arrivée sur la parcelle.

Ainsi, pour une même parcelle, plusieurs types de questionnaires ont pu être établis. Pour chacun des cas, les mêmes informations sur la parcelle ont été demandées : ses modifications par rapport aux bâtiments, à sa situation à l'égard du lotissement, à ses équipements.

Adéquation aux différentes situations

Au terme du deuxième passage de l'enquête et sur l'ensemble des thèmes évoqués, 213 questions ont été posées. Toutes ne concernent pas le ménage interviewé puisqu'elles dépendent de leur appartenance à une situation (exemple : propriétaire ou locataire). Malgré les 17 pages pour le premier passage, il faut relativiser la longueur du questionnaire en regard de la situation des personnes interrogées. Nous avons tenté de donner une image suffisamment exhaustive des divers cas de figure pouvant être rencontrés. Aussi un ménage pouvait répondre entre 85 et 164 questions pour le premier passage et entre 12 et 83 questions pour le deuxième.

Questionnaire du 1er passage

ORSTOM/IRSSH
BURKINA FASO

LES ENJEUX DES EXTENSIONS URBAINES A OUAGADOUGOU
1986 - 1987

FRANCE 1

CODAGE

J H A
Date : ⌊_⌋⌊_⌋⌊_⌋⌊_⌋

coordonnée du tirage ⌊_⌋⌊_⌋⌊_⌋⌊_⌋⌊_⌋⌊_⌋

lieu : secteur ⌊_⌋ sous-secteur ⌊_⌋⌊_⌋⌊_⌋
quartier : ..... ⌊_⌋⌊_⌋

Nombre de ménages vivant sur la parcelle : ⌊_⌋⌊_⌋

Nombre de personnes vivant sur la parcelle : ⌊_⌋⌊_⌋

IDENTIFICATION DU CM : Propriétaire/chef de concession/Locataire/hébergé/ ⌊_⌋

sexe : H ou F ⌊_⌋

Année de naissance : 19⌊_⌋⌊_⌋ Age : ⌊_⌋⌊_⌋

1. Marié (e) - 2. Divorcé (e) - 3. Veuf (ve) - 4. Célibataire ⌊_⌋

depuis combien de temps vivez-vous ici : date ⌊_⌋⌊_⌋ ⌊_⌋⌊_⌋
mois année

nombre d'années : ⌊_⌋⌊_⌋

LIEU DE NAISSANCE DU CM

Ouaga : secteur : ⌊_⌋ sous-secteur : ⌊_⌋⌊_⌋⌊_⌋⌊_⌋⌊_⌋⌊_⌋
quartier : .....

Ailleurs : Province .....
Département .....
Pays étranger.....
Ville /village.....

Ethnie ..... sous-groupe : ..... ⌊_⌋⌊_⌋ ⌊_⌋⌊_⌋
Nationalité : ..... ⌊_⌋

CODAGE 2

LIEU DE NAISSANCE DU PERE DU CM

Ouaga : secteur ⌊_⌋
quartier .....

Ailleurs : Province .....
Département.....
Pays étranger .....
Ville/village.....

Ethnie : ..... sous-groupe : ..... ⌊_⌋⌊_⌋ ⌊_⌋⌊_⌋
Nationalité : ..... ⌊_⌋

LIEU DE NAISSANCE DE LA MERE DU CM

Ouaga : secteur ⌊_⌋
quartier .....

Ailleurs : Province ..... ⌊_⌋⌊_⌋⌊_⌋
Pays étranger.....
Ville/village.....

Ethnie : ..... sous-groupe : ..... ⌊_⌋⌊_⌋ ⌊_⌋⌊_⌋
Nationalité :..... ⌊_⌋

COMPOSITION DU MENAGE DU CM

. Pour CM homme :

nombre d'épouses ⌊_⌋ ⌊_⌋

nombre d'épouses résidant ailleurs ⌊_⌋ ⌊_⌋

. pour les épouses résidant avec le CM

épouses	ETHNIE	LIEU DE NAISSANCE			Activité principale						Activité secondaire
		Ouaga province	Hors Ouaga ville/village	Hors Ouaga province	NATURE	Lieux: sur la parcelle	aux abords de la parcelle	sect de Ouaga	Faubourg	autre	nature
1											
	⌊_⌋		⌊_⌋		⌊_⌋			⌊_⌋			⌊_⌋
2											
	⌊_⌋		⌊_⌋		⌊_⌋			⌊_⌋			⌊_⌋
3											
	⌊_⌋		⌊_⌋		⌊_⌋			⌊_⌋			⌊_⌋
4											
	⌊_⌋		⌊_⌋		⌊_⌋			⌊_⌋			⌊_⌋

Lieux de résidence et motifs de la non résidence chez le CM pour les épouses résidant ailleurs :

4.

épouses	LIEUX DE RESIDENCE: Ouaga secteur	Ouaga quartier	hors Ouaga ville/village	hors Ouaga province	nbre d'enfants du CM résidant avec l'épouse	MOTIFS DE LA RESIDENCE AILLEURS: manque de place	mésentente des coépouses	occupation des terrains du CM	mésentente avec son mari	activité professionnelle	autres * à préciser	STATUT RESIDENTIEL: sur une autre parcelle du CM	hébergé chez parents du CM	hébergé chez des parents	locataire	autre	Activité
1	...	...	...	...	...												
			⌊_⌋		⌊_⌋				⌊_⌋					⌊_⌋			⌊_⌋
2	...	...	...	...	...												
			⌊_⌋		⌊_⌋				⌊_⌋					⌊_⌋			⌊_⌋
3	...	...	...	...	...												
			⌊_⌋		⌊_⌋				⌊_⌋					⌊_⌋			⌊_⌋
4	...	...	...	...	...												
			⌊_⌋		⌊_⌋				⌊_⌋					⌊_⌋			⌊_⌋

* autres motifs à préciser :.....

.-Pour CM femme :

CODAGE

. nombre de mariages successifs ⌊_⌋ ⌊_⌋

. si CM mariée : le mari a-t-il d'autres femmes ? oui/non ⌊_⌋

. motifs de la résidence séparée: (entourer la bonne réponse) : manque de place ; mésentents avec les autres épouses ; mésentente avec le mari ; occupation des terrains du CM ; activité professionnelle ; autre (à préciser) :..... ⌊_⌋

.-Pour tous les CM : (hommes ou femmes)

. nombre d'enfants du CM résidant dans la concession ⌊_⌋ ⌊_⌋

. nombre d'enfants du CM résidant ailleurs ⌊_⌋ ⌊_⌋

. Enfants du CM (résidents et non résidents)

Enfants	Sexe	Age	Marié (e) (M)	lieu de résidence : chez le CM/ Ouaga - secteur / province / étranger	Activité	lieu d'activité
1	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
2	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
3	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
4	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
5	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
6	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
7	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
8	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
9	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴
10	⌴	⌴⌴	⌴	⌴⌴⌴	⌴⌴	⌴⌴⌴

Autre personne (célibataire ou veuve sans le conjoint) dépendant du CM et vivant chez lui (depuis 3 mois ou plus)

personne	Sexe	Age	Marié (e)	Parenté avec le CM/ Parenté avec le conjoint/ autre lien .	Raison de l'installation chez le CM: Recherche d'emploi ; expulsion ; étude; employé par le CM ; manque de place ailleurs ; raison de santé ; autre *	hébergé = 1 locataire= 2	Activités
1	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
2	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
3	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
4	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
5	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
6	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
7	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴
8	⌴	⌴⌴	⌴	⌴	⌴	⌴	⌴⌴

* Autre = ...

ETAPES DE LA VIE DU CM

Etapes	Lieux de résidence successifs: Ouaga-secteur/province/étranger ville / village V v	Héberg H locataire L propriétaire P	Activités	Durée du séjour	Raisons du changement de résidence: Mariage	Volonté d'indépendance	recherche emploi	Améliorer les conditions de logement	accéder à la propriété	scolariser les enfants	déplacement professionnel	Accéder à un logement moins cher	Autres*
1	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
2	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
3	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
4	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
5	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
6	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
7	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					
8	⌊_⌊_⌊	⌊_⌋	⌊_⌋	⌊_⌋				⌊_⌋					

Autre* : .....

Nombre d'étapes : ⌊_⌋

9

FORMATION DU C.M.

PARTIE RESERVEE AU CODAGE

Ecole : .....Si oui, niveau de scolarisation..... ⌊_⌊_⌋

Ecole Professionnelle : si oui, laquelle ..... ⌊_⌊_⌋

Apprentissage : oui/non Formation sur le tas : oui /non..... ⌊_⌋

Ecole coranique : si oui, nombre d'années : ..... ⌊_⌊_⌋

Langues parlées : langue de son ethnie ; Français ; Moré ; Arabe ; autres = ⌊_⌋

Sait lire : oui/non ; sait écrire : oui/non..... ⌊_⌋

Religion : Musulman ; animiste ; catholique : protestant ; autres ; ⌊_⌋

ACTIVITES PROFESSIONNELLES DU C.M.

ACTIVITE PRINCIPALE : ..... ⌊_⌋

Nom et nature de l'entreprise : ..... ⌊_⌋

Le C.M. exerce-t-il actuellement cette activité ? oui/non

Si non -- Depuis combien de temps a-t-il cessé d'exercer cette activité ? ⌊_⌊_⌋

Nombre de mois /_/ Nombre d'années /_/

-- Pour quelles raisons a-t-il cessé cette activité : (entourer la bonne réponse) : ⌊_⌋

licenciement ; démission ; faillite de l'entreprise ; changement de résidence ; déguerpissement ; vieillesse, mise à la retraite ; mévente des productions ; incapacité physique ; fin de contrat ; succession familiale ; autre ;

-STATUT dans cette activité (entourer la bonne réponse) : 1 travailleur indépendant ; 2 aide - familial ; 3 manoeuvre occasionnel ; 4 travailleur sous contrat temporaire ou saisonnier ; 5 manoeuvre permanent ; 6 ouvrier ; 7 ouvrier qualifié ; 8 chef d'équipe ; 9 agent de maîtrise ; 10 employé ; 11 technicien ; 12 cadre moyen ; 13 cadre supérieur ; 14 petit patron (1 à 5 salariés) ; 15 patron (5 à 10 salariés) ; 16 entrepreneur (+ de 10 salariés ; préciser leur nombre : . . . . .) ⌊_⌋⌊_⌋

autre : . . . . . .

-LIEU D'EXERCICE de cette activité :.à OUAGA : Secteur /_/_/

Quartier..... Sous-secteur /_/_/

.En dehors de OUAGA : Province ..... ⌊_⌋⌊_⌋

Ville..... Village : oui/non

-DUREE DANS L'ACTIVITE: mois /_/_/ années /_/_/..... ⌊_⌋⌊_⌋

-GAINS DANS L'ACTIVITE: Par jour .....

Par mois ..... ⌊_⌋⌊_⌋⌊_⌋ J M

•DEUXIEME ACTIVITE : ..... ⌊_⌋

-STATUT dans cette activité (se reporter à l'énumération précédente) :..... ⌊_⌋

.....

-LIEU D'EXERCICE DE CETTE ACTIVITE :.OUAGA : Secteur /_/_/

Quartier.....

.En dehors de OUAGA : Province..... ⌊_⌋⌊_⌋

Ville ..... Village : oui/non

-DUREE DANS L'ACTIVITE: mois /_/ années /_/_/ ⌊_⌋⌊_⌋

-GAINS DANS L'ACTIVITE: Par jour .....

Par mois ..... ⌊_⌋⌊_⌋⌊_⌋ J M

•TROISIEME ACTIVITE : ..... ⌊_⌋

-STATUT dans cette activité (se reporter à l'énumération de la 1ère activité) :..... ⌊_⌋

.....

-LIEU D'EXERCICE DE CETTE ACTIVITE: :.OUAGA : Secteur /_/_/

Quartier .....

.En dehors de OUAGA : Province ..... ⌊_⌋⌊_⌋

Ville.....Village : oui/non

•ACTIVITE DU PERE DU C.M. ..... ⌊_⌋

-STATUT DANS L'ACTIVITE (se reporter à l'énumération plus haut) : .....

..... ⌊_⌋

-LIEU D'EXERCICE :.OUAGA : Secteur /_/_/

Quartier .....

.En dehors de OUAGA : Province ..... Ville ..... ⌊_⌋⌊_⌋

Pays..... Village : oui/non

STATUT FONCIER

-Etes-vous propriétaire de votre parcelle ? oui/non ⌊_⌋

Si oui: avez-vous un titre foncier ? oui/non. Si oui, date = ... ⌊_⌋⌊_⌋

Avez-vous un permis urbain d'habiter ? oui/non. Si oui, date =..... ⌊_⌋⌊_⌋

Comment avez-vous obtenu cette parcelle ?

-Par attribution par l'Etat ? Si oui, date :..... ⌊_⌊_⌊_⌋

.Cette attribution a-t-elle fait suite à un déguerpissement ?

.Si oui, date du déguerpissement..... Nom du quartier déguerpi..... ⌊_⌊_⌊_⌊_⌋

.Somme déjà payée pour la parcelle attribuée :..... Somme à payer..... ⌊_⌊_⌊_⌋ ⌊_⌊_⌊_⌋

-Par achat ou par attribution coutumière ? Date :..... Dépenses : argent.....autre..... ⌊_⌊_⌊_⌋ ⌊_⌊_⌊_⌋

.Le vendeur était = un chef de terre ? un chef de village ? un chef de quartier ?
un particulier ? autre =..... ⌊_⌋

.Etait-il un parent ? Si oui, lien de parenté :..... ⌊_⌋

-Par héritage ? Si oui, date :..... ⌊_⌊_⌋

.Vous avez hérité : du père ? d'un frère du père ? d'un frère de la mère ? de la mère ?
d'un frère de même père ? d'un frère de même mère (mais de père différent) ? du mari ? autre= ⌊_⌋

.Si un autre parent, précisez la parenté .....

-Par un don ? Si oui, date ..... ⌊_⌊_⌋

.Le donateur était-il un parent ? oui / non

.Si oui précisez le lien de parenté : du père ? d'un frère du père ? d'un frère de la mère ?
de la mère ? d'un frère de même père ? d'un frère de même mère (mais de père différent) ? ⌊_⌊_⌋
du mari ? autre = ...

.Si un autre parent, précisez la parenté.....

.Raison du don : possède plusieurs parcelles ; pour ne pas être exproprié ; incapacité de
construire ; pour loger les enfants ; autres:..... ⌊_⌋

- Quelle est la situation de votre parcelle par rapport au lotissement ? (possibilité de plusieurs réponses) ⌊_⌊_⌊_⌋

.Pas encore lotie ; le bornage est en cours ; le bornage est terminé ; les limites de la parcelle sont modifiées ; ma parcelle est sur une réserve administrative ; ma parcelle est sur l'emplacement de la future voirie ; une autre parcelle ne m'a pas encore été attribuée ; une autre parcelle m'a été attribuée dans le quartier ; dans un autre quartier

.Si autre quartier, précisez lequel ..... ⌊_⌋

.Avez-vous reçu une indemnité pour le déguerpissement ? oui / non ⌊_⌊_⌋

Si oui, combien avez-vous perçu ?.....

14

ETES VOUS LOCATAIRE ? oui/non ⌊_⌋

Si oui depuis combien de temps ? mois ⌊_⌋ années ⌊_⌋ ⌊_⌋

Est-ce à la suite d'un déguerpissement ? oui/non ⌊_⌋

Louez-vous toute la parcelle ? oui/non. Nombre de pièces louées ⌊_⌋ ⌊_⌋

Montant total de la location mensuelle : ..... ⌊_⌋

Combien de temps comptez-vous rester ici encore ? mois ⌊_⌋ année ⌊_⌋ ⌊_⌋

ETES-VOUS HEBERGE ? oui/non ⌊_⌋

Si oui depuis combien de temps ? mois ⌊_⌋ Année ⌊_⌋ ⌊_⌋

Est-ce à la suite d'un déguerpissement ? oui/non ⌊_⌋

Participez vous à quelques dépenses ? oui/non ⌊_⌋

Combien de temps comptez-vous rester ici encore ? mois ⌊_⌋ année ⌊_⌋ ⌊_⌋

SI VOUS ETES LOCATAIRE OU HEBERGE :

. Lieu de résidence du propriétaire non résident :
Ouaga-secteur ⌊_⌋ hors Ouaga-Province - - - - - - - - - - - - - - ⌊_⌋ ⌊_⌋

. Quelle est mon activité ? .....---------- - - - - - - - - ⌊_⌋ ⌊_⌋

. Etes-vous parent avec le propriétaire ? oui / non ⌊_⌋

. si oui : le propriétaire est-il :

un parent paternel : père/frère du père/frère aîné/soeur du père/soeur/autre

un parent maternel : mère/frère de la mère/soeur de la mère/autre

un parent par alliance : mari/épouse/père du conjoint/mère du conjoint/autre ⌊_⌋ ⌊_⌋

un "enfant" : fils/fille/fils de fils/fils de fille/fils de frère/fils de soeur/autre

Autres : précisez . . . . . .

15

ETES VOUS RELOGE PROVISOIREMENT PAR L'ETAT ? oui/non CODAGE

si oui, depuis combien de temps ? mois ⌊_⌋ années ⌊_⌋ ⌊_⌋ ⌊_⌋

POUR TOUS (PROPRIETAIRES, LOCATAIRES, HEBERGES OU RELOGES) :

- Possédez-vous d'autres parcelles d'habitation ? oui/non ⌊_⌋ ⌊_⌋
si oui combien : . . . . . .

- Emplacement de ces parcelles :

	Ouaga secteur	Date d'acquisition ou d'obtention	hors Ouaga province	Date	
1	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋ ⌊_⌋
2	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋ ⌊_⌋
3	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋ ⌊_⌋

- êtes-vous propriétaire de champs de culture ? oui/non si oui combien : ⌊_⌋ ⌊_⌋
êtes-vous locataire de champs de culture ? oui/non si oui combien : ⌊_⌋ ⌊_⌋

- emplacement des champs de culture :

	champs possédés		champs loués		
	Ouaga-secteur	hors ouaga-province	Ouaga-secteur	hors ouaga-province	
1	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋
2	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋
3	.....	.....	.....	.....	⌊_⌋ ⌊_⌋ ⌊_⌋

- DESCRIPTION DE LA PARCELLE

Longueur : ⌊__⌋ m largeur : ⌊__⌋ m ⌊__⌋

Nombre de bâtiments à usage d'habitation ⌊__⌋ ⌊__⌋

Nombre de pièces à usage d'habitation ⌊__⌋ ⌊__⌋

Equipements :

Puits	forage	eau courante	borne fontaine dans environs	électricité	groupe électrogène	télévision	W.C.	Vélo	Deux roues moteur	automobile	bétail	poulailler	jardin potager	arbres	atelier ou boutique dans concession	autres
⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋

Autre (à préciser) : .....

17.

Construction et mode de financement des bâtiments :

N° bâtiment	Nbre de pièces	Date de construction	matériaux : mur	matériaux : toiture	coûts estimés	mode de construction : CM seul = 1, CM+tâcheron = 2, CM+MO familiale = 3, tâcheron seul = 4, autre = 9	mode de financement : épargne personnelle = 1	épargne familiale = 2	Tontine = 3	emprunt à la banque = 4	emprunt à un employeur = 5	emprunt à un particulier	* Autre
1	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
2	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
3	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
4	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
5	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
6	⌊__⌋	⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋
clôture		⌊__⌋	⌊__⌋		⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋	⌊__⌋

* si autre précisez = ...

Questionnaire du 2ème passage : nouveau ménage arrivé sur la parcelle

O G R S T	LES ENJEUX DES EXTENSIONS URBAINES A OUAGADOUGOU	O R S T O M
(RSSH)	- 2ème passage 1987 -	Département D
BURKINA FASO	(Questionnaire type A)	FRANCE

J M A

1. DATE |__|__|__|__|__|__| n° du ménage : |__|__|

. COORDONNEES TIRAGE |__|__|__|__|__|__|

2. LIEU SECTEUR : QUARTIER : |__|__|__|__|

3. NOMBRE DE MENAGES VIVANT SUR LA PARCELLE : |__|__|

4. NOMBRE DE PERSONNES VIVANT SUR LA PARCELLE : |__|__|

5. NOMBRE DE PERSONNES VIVANT DANS LE MENAGE : |__|__|

IDENTIFICATION DU CM

6. STATUT D'OCCUPATION : |__|__|

7. SEXE : . AGE : |__| |__|__|

8. STATUT MATRIMONIAL : |__|__|

9. ANCIENNETE SUR LA PARCELLE : |__|__|__|

LIEU DE NAISSANCE DU CM

10. Ouaga : secteur : |__|__| sous-secteur : |__|__|
quartier : .....

11. Ailleurs : Province .....
Département .....
Pays étranger.....
Ville /village.....

|__|__|__|__|__|__|

12. Ethnie ..... sous-groupe : ..... |__|__| |__|__|

13. Nationalité : ..... |__|__|

COMPOSITION DU MENAGE

2.

- Pour un homme

14 - Nombre d'épouses présentes |__|__|

15 - Nombre d'épouses absentes |__|__|

- Pour CM femme :

16. nombre de mariages successifs |__|__|

17. si CM mariée : le mari a-t-il d'autres femmes ? oui/non |__|__|

18. motifs de la résidence séparée : (entourer la bonne réponse) : manque de place ; mésentente avec les autres épouses ; mésentente avec le mari ; occupation des terrains du CM ; activité professionnelle ; autre (à préciser) :..... |__|__|

- Pour tous les CM : (hommes ou femmes)

19. nombre d'enfants du CM résidant dans la concession |__|__| |__|__|

20. nombre d'enfants du CM résidant ailleurs |__|__| |__|__|

21 - Nombre de dépendants |__|__|

22 Dépendants du CM :

Personnes	Sexe	Age	Parenté avec le CM / Parenté avec le conjoint / autre lien	Activité
1				
2				
3				
4				
5				
6				

ACTIVITES PROFESSIONNELLES DU C.M.

23. ACTIVITE PRINCIPALE : .... ⌊_⌋⌊_⌋

24. Nom et nature de l'entreprise : ..... ⌊_⌋

25. STATUT dans cette activité (entourer la bonne réponse) : 1 travailleur indépendant ; 2 aide - familial ; 3 manoeuvre occasionnel ; 4 travailleur sous contrat temporaire ou saisonnier ; 5 manoeuvre permanent ; 6 ouvrier ; 7 ouvrier qualifié ; 8 chef d'équipe ; 9 agent de maîtrise ; 10 employé ; 11 technicien ; 12 cadre moyen ; 13 cadre supérieur ; 14 petit patron (1 à 5 salariés) ; 15 patron (5 à 10 salariés) ; 16 entrepreneur (+ de 10 salariés ; préciser leur nombre : . . . . . .)

⌊_⌋⌊_⌋

autre : . . . .

26. LIEU D'EXERCICE de cette activité : à OUAGA : Secteur ⌊_⌋⌊_⌋

Quartier..... Sous-secteur ⌊_⌋⌊_⌋

En dehors de OUAGA : Province .....

⌊_⌋⌊_⌋

Ville..... Village : oui/non

27. DEPUIS DANS L'ACTIVITE mois ⌊_⌋⌊_⌋ année ⌊_⌋⌊_⌋..... ⌊_⌋⌊_⌋

ETAPE PRECEDENTE PRECEDANT L'INSTALLATION SUR LA PARCELLE

29 - Lieu de résidence : ⌊_⌋⌊_⌋⌊_⌋

29 - Statut d'occupation : ⌊_⌋

30 - Activité : ⌊_⌋⌊_⌋

31 - Durée de l'étape : ⌊_⌋⌊_⌋

32 - Motif de départ de la parcelle précédente : ..... ⌊_⌋

.....

33 - Motif d'installation sur la parcelle actuelle : ..... ⌊_⌋

.....

STATUT FONCIER

34. Etes-vous propriétaire de votre parcelle ? oui/non ⌊_⌋

35 Si oui: avez-vous un titre foncier ? oui/non. Si oui, date = ... ⌊_⌋⌊_⌋

36 Avez-vous un permis urbain d'habiter ? oui/non. Si oui, date =..... ⌊_⌋⌊_⌋

Avez-vous une attestation d'attribution. Si oui date : .....

Comment avez-vous obtenu cette parcelle ?

.Par attribution par l'Etat ? Si oui, date :..... ....

.Cette attribution a-t-elle fait suite à un déguerpissement ?

.Si oui, date du déguerpissement..... Nom du quartier déguerpi.....

.Somme déjà payée pour la parcelle attribuée :..... Somme restant à payer.....

.Par achat ou par attribution coutumière ? Date :..... Dépenses : argent.....autre.....

.Le vendeur était = un chef de terre ? un chef de village ? un chef de quartier ?
un particulier ? autre =.....

.Etait-il un parent ? Si oui, lien de parenté :.....

.Par héritage ? Si oui, date :.....

.Vous avez hérité : du père ? d'un frère du père ? d'un frère de la mère ? de la mère ?
d'un frère de même père ? d'un frère de même mère (mais de père différent) ? du mari ? autre=
.Si un autre parent, précisez la parenté .....

.Par un don ? Si oui, date .....

.Le donateur était-il un parent ? oui / non

.Si oui précisez le lien de parenté : du père ? d'un frère du père ? d'un frère de la mère ?
de la mère ? d'un frère de même père ? d'un frère de même mère (mais de père différent) ?
du mari ? autre = ...

.Si un autre parent, précisez la parenté.....

.Raison du don : possède plusieurs parcelles ; pour ne pas être exproprié ; incapacité de
construire ; pour loger les enfants ; autres:.....

.ETES VOUS LOCATAIRE ? oui/non

.Si oui depuis combien de temps ? mois ___ années ___

.Est-ce à la suite d'un déguerpissement ? oui/non Qui paie le loyer ? .....

.Louez-vous toute la parcelle ? oui/non. Nombre de pièces louées ___

.Montant total de la location mensuelle : .....

Combien de temps comptez-vous rester ici encore ? mois ___ année ___

.ETES-VOUS HEBERGE ? oui/non

.Si oui depuis combien de temps ? mois ___ Année ___

.Est-ce à la suite d'un déguerpissement ? oui/non

.Participez vous à quelques dépenses ? oui/non : Lesquelles ? .....

.Combien de temps comptez-vous rester ici encore ? mois ___ année ___

SI VOUS ETES LOCATAIRE OU HEBERGE :

.Lieu de résidence du propriétaire non résident :
Ouaga-secteur ___ hors Ouaga-Province - - - - - - - - - - - - - -

.Quelle est son activité ? .....---------- - - - - - - - - -

.Etes-vous parent avec le propriétaire ? oui / non

.si oui : le propriétaire est-il :

un parent paternel : père/frère du père/frère aîné/soeur du père/soeur/autre

un parent maternel : mère/frère de la mère/soeur de la mère/autre

un parent par alliance : mari/épouse/père du conjoint/mère du conjoint/autre

un "enfant" : fils/fille/fils de fils/fils de fille/fils de frère/fils de soeur/autre

Autres : précisez . . . . . .

46. EVOLUTION DE LA SITUATION DE LA PARCELLE PAR RAPPORT AUX OPERATIONS DE DE LOTISSEMENT

.....

.....

.....

.....

.....

DESCRIPTION DE LA PARCELLE

47. MODIFICATIONS DE LA SUPERFICIE DEPUIS : AN : |__|__|

47. CONSTRUCTION DE NOUVEAUX BATIMENTS :

N° BATIMENT	NBRE DE PIECES	DATE DE CONSTRUCTION	Matériaux MUR	Matériaux TOITURE	COUTS ESTIMES	MODE DE CONSTRUCTION	EPARGNE PERSONNELLE = 1	EPARGNE FAMILIALE = 2	TONTINE = 3	EMPRUNT A LA BANQUE = 4	EMPRUNT A UN EMPLOYEUR = 5	EMPRUNT A UN PARTICULIER = 6	Autre = 9
a	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|
b	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|
c	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|

48. NOUVEAUX EQUIPEMENTS ET DISPARITION D'EQUIPEMENTS OBSERVES EN 1986

PUITS	FORAGE	EAU COURANTE	BORNE FONTAINE DANS ENVIRONS	ELECTRICITE	W.C.	ARBRES	AUTRES EQUIP. PARCELLE	GROUPE ELECTROGENE	TELEVISION	VELO	VELOMOTEUR, MOTO	AUTOMOBILE	BETAIL, POULAILLER	JARDIN POTAGER	ATELIER OU BOUTIQUE	AUTRE EQUIP. MENAGE
|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|

Questionnaire du 2ème passage : ménage enquêté au 1er passage

DGRST
(IRSSH)
BURKINA FASO

LES ENJEUX DES EXTENSIONS URBAINES A OUAGADOUGOU
- 2ème passage 1987 -
(Questionnaire type B)

ORSTOM
Département D
FRANCE

J M A

1. DATE |__|__|__|__|__|__| |__|__|

2. COORDONNEES TIRAGE |__|__|__|__|__|__|

3. LIEU SECTEUR : QUARTIER : |__|__|__|__|

4. NOMBRE DE MENAGES VIVANT SUR LA PARCELLE : |__|__|

5. NOMBRE DE PERSONNES VIVANT SUR LA PARCELLE : |__|__|

6. NOMBRE DE PERSONNES VIVANT DANS LE MENAGE : |__|__|

IDENTIFICATION DU CM

7. CHANGEMENT DU STATUT D'OCCUPATION NATURE : |__|__|
DATE : |__|__|

8. CHANGEMENT DU STATUT MATRIMONIAL NATURE : |__|__|
DATE : |__|__|

COMPOSITION DU MENAGE

9. EPOUSES PRESENTES EN 86 AYANT QUITTE CONCESSION (dates) : |__|__|

10. EPOUSES ABSENTES EN 86 ET RESIDENTES EN 87 (dates) : |__|__|

NOUVELLES EPOUSES PRESENTES

	Ethnie	Lieu naissance	Résidence antérieure	Activité
1	|__|__|	|__|__|__|	|__|__|__|	|__|__|
2	|__|__|	|__|__|__|	|__|__|__|	|__|__|
3	|__|__|	|__|__|__|	|__|__|__|	|__|__|

EPOUSES ABSENTES

	Nouvelle résidence	Motif du départ
1	|__|__|__|	|__|__|
2	|__|__|__|	|__|__|
3	|__|__|__|	|__|__|

11. POUR LES CM FEMMES
SE SONT-ELLES MARIEES DEPUIS 1 AN ? |__|__|
CHANGEMENT DE STATUT PAR RAPPORT AU MARI ? |__|__|

12. POUR TOUS CM ENFANTS NES DEPUIS 1 AN : |__|__|

13. ENFANTS ABSENTS EN 86 ET RESIDENTS EN 87 : |__|__|

14. ENFANTS PRESENTS EN 86 ET ABSENTS EN 87 : |__|__|

Enfants	Sexe	Age	Marié (e) (M)	Nouveau lieu de résidence :	Activité	Motif départ ou Année.
1	|__|	|__|__|	|__|	|__|__|__|	|__|__|	|__|__|__|
2	|__|	|__|__|	|__|	|__|__|__|	|__|__|	|__|__|__|
3	|__|	|__|__|	|__|	|__|__|__|	|__|__|	|__|__|__|
4	|__|	|__|__|	|__|	|__|__|__|	|__|__|	|__|__|__|
5	|__|	|__|__|	|__|	|__|__|__|	|__|__|	|__|__|__|

OBSERVATIONS :

15. DEPENDANTS ARRIVES DEPUIS 1 AN |__|__|

16. DEPENDANTS PARTIS AU COURS DE L'ANNEE ECOULEE |__|__|

Personne	Sexe	Age	Marié (e)	Parenté avec le CM/ Parente avec le conjoint/ autre Lien	Raison de l'installation chez le CM ou du départ	Hébergé : 1 Locataire : 2 ou nouvelle résidence	Activités
1	|__|	|__|__|	|__|	|__|__|	|__|	|__|__|
2	|__|	|__|__|	|__|	|__|__|	|__|	|__|__|
3	|__|	|__|__|	|__|	|__|__|	|__|	|__|__|
4	|__|	|__|__|	|__|	|__|__|	|__|	|__|__|
5	|__|	|__|__|	|__|	|__|__|	|__|	|__|__|

FORMATION DU CM

17. NOUVELLES FORMATIONS ENTREPRISES AU COURS DE L'ANNEE ECOULEE |__|__|

ACTIVITES DU CM

18. CHANGEMENTS DE L'ACTIVITE PRINCIPALE DEPUIS 1 AN

NATURE DE L'ACTIVITE : |__|__|

ETABLISSEMENT : |__|__|

STATUT DANS L'ACTIVITE : |__|__|

STATUT FONCIER

19. ACCESSION A LA PROPRIETE DEPUIS 1 AN

	Au lieu de résidence	Ailleurs
PAIE	|__|__|	|__|__|
MENSUELLES	|__|__|	|__|__|
COUT	|__|__|	|__|__|

20. POUR LES CHEFS DE CONCESSION NON PROPRIETAIRES :

- MODIFICATIONS DE L'ESPACE OCCUPE DEPUIS 1 AN ..... |__|__|
- MODIFICATIONS DU PRIX DE LOCATION OU DES CONDITIONS D'HEBERGEMENT ..... |__|__|
- CHANGEMENT DE PROPRIETAIRE ..... |__|__|

21. EVOLUTION DE LA SITUATION DE LA PARCELLE PAR RAPPORT AUX OPERATIONS DE DE LOTISSEMENT

.....

.....

.....

.....

.....

DESCRIPTION DE LA PARCELLE

22. MODIFICATIONS DE LA SUPERFICIE DEPUIS 1 AN : |__|__|

23. CONSTRUCTION DE NOUVEAUX BATIMENTS :

N° BATIMENT	NBRE DE PIECES	DATE DE CONSTRUCTION	Matériaux MUR	Matériaux TOITURE	COUTS ESTIMES	MODE DE CONSTRUCTION	EPARGNE PERSONNELLE : 1	EPARGNE FAMILIALE : 2	TONTINE : 3	EMPRUNT A LA BANQUE : 4	EMPRUNT A UN EMPLOYEUR : 5	EMPRUNT A UN PARTICULIER : 6	AUTRES : 9
1	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|
2	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|
3	|__|	|__|__|	|__|	|__|	|__|__|__|__|	|__|	|__|	|__|	|__|	|__|	|__|	|__|

24. NOUVEAUX EQUIPEMENTS ET DISPARITION D'EQUIPEMENTS OBSERVES EN 1986

	PUITS	FORAGE	EAU COURANTE	BORNE FONTAINE DANS ENVIRONS	ELECTRICITE	W.C.	ARBRES	AUTRES EQUIP. PARCELLE	GROUPE ELECTROGENE	TELEVISION	VELO	VELOMOTEUR, MOTO	AUTOMOBILE	BETAIL, POULAILLER	JARDIN POTAGER	ATELIER OU BOUTIQUE	AUTRE EQUIP. MENAGE
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Questionnaire du 2ème passage : ménage parti avant le 2ème passage

D G R S T	LES ENJEUX DES EXTENSIONS URBAINES A OUAGADOUGOU	O R S T O M
(IRSSH)	- 2ème passage 1987 -	Département D
BURKINA FASO	(Questionnaire type 2)	FRANCE

1 . DATE :	|__|__|___|___|___|
2 . COORDONNEES DU TIRAGE	|__|__|___|___|
3 . LIEU SECTEUR :	
QUARTIER :	|___|__|
4 . NOMBRE DE MENAGES VIVANT SUR PARCELLE	|___|__|
RAPPEL 1er PASSAGE :	|___|__|
NOMBRE DE PERSONNES VIVANT DANS LE MENAGE	|___|__|
NOMBRE DE PERSONNES VIVANT SUR PARCELLE	|___|__|
RAPPEL 1er PASSAGE	|___|__|
IDENTIFICATION DU CM AYANT QUITTE LA PARCELLE	
5 . STATUT D'OCCUPATION	|___|__|
6 . SEXE : 7. AGE :	|__| |___|__|
8 . STATUT MATRIMONIAL	|___|__|
9 . ANCIENNETE SUR LA PARCELLE	|__|__|__|
10 DATE DU DEPART DE LA PARCELLE	|__|__|

OBSERVATIONS :

11 . MOTIFS DU DEPART	|__|__|__|
.....	
.....	
.....	
.....	
12. NOUVELLE LOCALISATION DU MENAGE	|__|__|__|
POUR OUAGA SECTEUR ET QUARTIER	
HORS OUAGA DEPARTEMENT ET LOCALITE	
13. NOUVEAU STATUT D'OCCUPATION DU SOL	|__|__|__|
14. NOUVEAU STATUT PROFESSIONNEL	|__|__|__|

Annexe 1.3 : Déroulement de l'enquête

Seule la connaissance des conditions de la collecte des données permet, dans notre situation, d'estimer la fiabilité des résultats.

Il importe, d'une part de répondre aux problèmes immédiats survenus au cours de l'enquête et, d'autre part d'être en mesure de cerner les difficultés rencontrées pouvant avoir des répercussions sur l'interprétation des résultats.

Le chronogramme (figure n°35) permet de situer cette phase dans l'exécution générale de l'enquête.

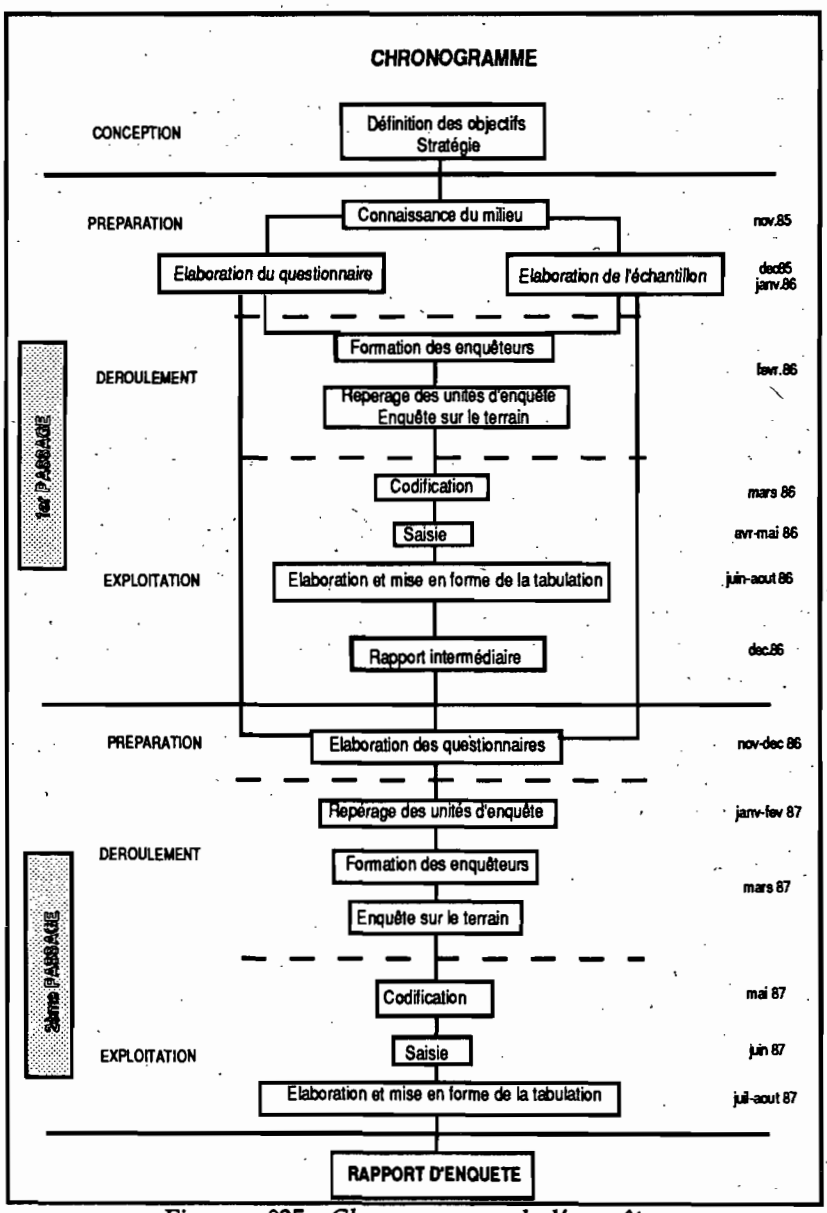


Figure n°35 : *Chronogramme de l'enquête*

Préparation de l'enquête sur le terrain

Phase de sensibilisation

Cette phase de 3 à 4 mois a débuté dès la préparation du questionnaire puis de l'échantillon. Une première démarche a été effectuée auprès des délégués CDR pour recueillir les données nécessaires à la constitution de la base de sondage. Afin d'officialiser cette démarche de sensibilisation, l'équipe est intervenue pour présenter cette enquête lors d'une réunion générale de tous les responsables CDR de chaque secteur. Ceux-ci ont par conséquent été sollicités pour informer les gens qu'une partie d'entre eux allait être sélectionnée pour participer à une enquête.

Nous avons insisté sur cette phase dans la mesure où, parallèlement à l'enquête, certains quartiers étaient touchés par la restructuration de l'espace périphérique et donc par l'attribution ou non de parcelles. Nous appréhendions alors d'une part une réticence de la part des enquêtés en fonction de cet aspect, la crainte d'être déguerpis et d'autre part une certaine saturation compte tenu des nombreuses enquêtes ou recensements ayant eu lieu ces dernières années.

Cette phase a été efficace puisque nous n'avons eu aucun refus de participation.

Calendrier de l'enquête sur le terrain

Pour les deux passages, l'enquête s'est déroulée sur une période de 21 jours consécutifs. Le premier passage a eu lieu entre le 1er et le 21 mars 1986 et le deuxième entre le 7 et le 28 février 1987.

Ce délai est minimum pour à la fois enquêter et repérer correctement les parcelles, en comptant une moyenne de quatre questionnaires par jour (on a estimé une durée variant entre une heure et trois heures pour repérer la parcelle et enquêter les chefs de ménages). Il est maximum pour que l'échantillon ne soit pas trop perturbé par les changements importants pouvant intervenir dans certaines zones menacées de restructuration. Le temps relativement réduit implique par conséquent un nombre important d'enquêteurs.

Les deux périodes d'enquête ne sont pas identiques puisque un mois sépare le début des deux passages. Mais ceci a une importance moindre pour notre enquête qui traite avant tout du foncier. Il s'agissait surtout de respecter la saison sèche.

Recrutement et formation des enquêteurs

12 enquêteurs ont été recrutés auprès de l'INSD, et encadrés par 6 superviseurs. Les exigences étaient de parler au moins le français et le *mooré* et d'avoir participé au recensement, cette dernière condition étant favorable pour le repérage des parcelles. Chaque enquêteur avait à charge entre deux et trois secteurs si possible contigus et pas trop éloignés de leur lieu de résidence afin de minimiser les déplacements, de pouvoir retourner facilement sur le lieu de l'enquête le cas échéant, et de bien connaître d'emblée les lieux.

Les enquêteurs ont reçu une formation de trois jours durant laquelle leur ont été fournies les instructions et les explications. Une journée de simulation a permis aux enquêteurs de tester et de comprendre le questionnaire et aux chercheurs de répondre à leurs propres questions sur celui-ci.

Identification des parcelles et des ménages

Premier passage

Chaque parcelle est identifiable par un numéro affecté par les agents recenseurs de l'INSD lors du recensement. Les enquêteurs recrutés possédaient déjà la logique de numérotation mais ont reçu, pour un terrain inconnu et si nécessaire, l'appui des CDR. Aussi le repérage des parcelles lors du premier passage n'a pas posé de problème majeur.

Celui-ci a été effectué presque simultanément avec les interviews, l'enquêteur, dans la mesure du possible, repérant la parcelle la veille de son passage ce qui lui permettait de prendre rendez-vous. Chaque jour le superviseur sélectionnait à l'intérieur de l'échantillon (qui, rappelons le, a été sur-évalué en prévision des parcelles détruites ou inhabitées) les parcelles à enquêter. Si une parcelle était détruite ou inoccupée, elle était remplacée par une autre parcelle toujours choisie dans l'échantillon. Si elle était apparemment habitée mais personne n'était en mesure de fournir les renseignements souhaités, l'enquêteur avait pour consigne d'y retourner régulièrement jusqu' au terme de l'enquête où il était décidé du tirage d'une autre parcelle.

Deuxième passage

Le repérage des parcelles lors du deuxième passage peut apparaître comme problématique en raison de l'anonymat de l'enquêté qu'il a été demandé de respecter. Comment alors repérer la parcelle dans un espace qui en une année a pu être considérablement modifié?

Ayant conscience de cette difficulté, nous avons demandé auparavant à l'INSD d'aviser les enquêteurs ayant participé au premier passage de coopérer au deuxième. Ainsi sur les douze enquêteurs, dix ont pu répondre à notre demande, les deux autres ayant pu participer au repérage des parcelles de leurs zones avec leurs remplaçants. En effet si lors du premier passage, enquête et repérage ont été simultanés, le repérage du deuxième passage a été effectué indépendamment et a eu lieu une semaine avant le début de l'enquête.

Ainsi sur les 853 parcelles, 13 n'ont pu être identifiées ce qui correspond à 1,5% de pertes, chiffre négligeable au regard des mouvements survenus en une année. Ce deuxième repérage a donc été très satisfaisant, l'enquêteur retrouvant assez bien les parcelles enquêtées du premier passage. Cependant, dans le cas où se posait le problème d'identification, nous avions recours à la connaissance du lieu de l'activité du chef de ménage.

Une autre difficulté s'est posée pour le deuxième passage de l'enquête; celle du ménage qui avait émigré. Ce n'est donc plus un problème d'identification de la parcelle, puisque les informations la concernant ont pu être observées par l'enquêteur, mais de renseignements portant sur ce ménage. Il s'est avéré que la plupart d'entre eux ont été déguerpis et relogés dans le même secteur. Ils ont donc pu être retrouvés avec la participation des CDR ou des voisins. Les informations recherchées ont été fournies, même sommairement, soit par le voisinage qui n'a pas montré de réticence à nous renseigner, soit par les ménages corésidents si le ménage partageait la parcelle.

Passages et contrôle

Chaque jour, le superviseur retrouvait ses enquêteurs, d'une part pour leur fournir la liste des parcelles à enquêter le jour suivant et faire un bilan des parcelles visitées, d'autre part pour reprendre chaque formulaire rempli afin d'éviter d'éventuelles erreurs ou omissions ne pouvant plus être rectifiées au moment de la codification.

Les premières semaines, surtout lors du premier passage, ont nécessité de nombreux retours sur le

terrain et c'est à peu près à ce terme que le questionnaire a été considéré comme assimilé. Ceci n'a pas exclu des retours jusqu'à la fin de l'enquête pour données manquantes ou incohérentes avec d'autres informations relevées dans le questionnaire.

Le deuxième passage a été intégré plus vite par les enquêteurs malgré l'identification des types de ménage qui fut l'une des principales difficultés de ce passage. Ceci a surtout concerné les ménages non migrants ayant subi des mouvements internes tels qu'un changement de chef de ménage ou un mariage d'un enfant ou dépendant résidant toujours sur la parcelle et créant, par la même, un nouveau ménage. Une autre difficulté fut de saisir les caractéristiques relatives au ménage émigré mais dont les caractéristiques et la situation de la parcelle correspondante ont été simples à évaluer sans l'intervention de l'enquêté (faire un état de la parcelle relève de l'observation et sa situation est semblable aux parcelles contigües). Les questionnaires correspondants n'ont été que partiellement remplis. De plus un contrôle plus strict a permis de vérifier la pertinence des réponses soit du premier soit du deuxième passage.

Les contrôles sur le terrain devaient être effectués par le superviseur à sa convenance suivant les difficultés particulières liées soit au secteur (bouleversements importants, manque de collaboration des responsables CDR), soit à l'enquêteur (difficultés pour le repérage), soit encore à l'enquêté (refus de répondre). Il faut préciser que les refus de participation ont été inexistants malgré certaines réticences. Quoiqu'il en soit, il était vivement recommandé au contrôleur d'effectuer quelques vérifications sur le travail de l'enquêteur.

Enfin les superviseurs devaient se réunir régulièrement pour confronter les divers problèmes rencontrés, permettant ainsi de mieux fixer dans une même logique les concepts de base.

Les biais

Le travail de terrain n'a pas posé de réels problèmes puisque ceux-ci, supposés concerner essentiellement l'élaboration de l'échantillon, ont pu être décelés avant le début de l'enquête. Mais les réponses à certaines questions n'ont pas toujours été à la mesure des informations souhaitées. Aussi il est nécessaire d'évaluer les biais éventuels dans les données afin d'éviter les risques d'erreurs au niveau de l'interprétation.

Pour cela nous distinguons deux types de biais:

- **les biais pré-identifiés** résultant de questions délicates qui sont source de suspicion de la part des enquêtés ou de notions qui ont été mal appréhendées lors de la préparation du questionnaire. Il s'agit :
  - des propriétaires qui ont, semble-t-il, été sur-représentés dans les déclarations. Est-ce dans une optique de stratégie foncière vis à vis d'une éventuelle attribution de parcelle accordée à cette catégorie?
  - de la possession d'autres parcelles et notamment à Ouagadougou qui, par contre, a été réellement dissimulée pour les mêmes raisons d'attribution de parcelle.
  - des activités multiples qui sont appréciées différemment selon les individus et des activités des épouses qui sont largement sous-représentées. Il n'est pas rare dans ce pays d'exercer plusieurs activités, dont l'une est souvent dans le secteur agricole (de façon occasionnelle ou semi-permanente) et il est notoire que les femmes déclarées inactives par le chef de ménage ont une activité dans le secteur informel.
  - des revenus qui ont été soit donnés par une réponse évasive ou non dits car il est difficile d'évaluer un revenu sur une question unique, soit sous-estimés ce qui est relativement fréquent. (Pour pallier à ce défaut il existe des indicateurs économiques liés pour la plupart aux dépenses mais cela n'a pas été une réelle préoccupation du projet).
  - du nombre de naissances et surtout de décès mesurés par le deuxième passage qui a également été sous-évalué comparativement aux études essentiellement démographiques.

Ces trois dernières questions, mal perçues par notre enquête, relèvent, pour avoir plus de

précision, d'enquêtes spécifiques. Nous avons dû nous contenter d'indicateurs sommaires issus de l'enquête.

- **les biais post-identifiés** issus d'une mauvaise formulation des questions ou d'une non compréhension de certains critères de base de la part des superviseurs. Il faut noter à cet effet que cela provient de concepts qui apparemment ne devaient pas poser de problèmes et qui n'ont donc jamais été explicites. Il s'agit :

  - du mode d'obtention de la parcelle pouvait faire l'objet de deux catégories de réponses non exclusives, aussi l'information aurait pu être plus complète qu'elle ne l'a été: il était tout à fait envisageable d'être attributaire d'une parcelle et de l'avoir initialement obtenue par un mode d'acquisition traditionnel. La question n'ayant pas été formulée dans ce sens, tous les enquêteurs n'ont pas insisté sur cet aspect. Cependant, les hébergés et surtout les locataires ont, pour certains, été dans l'incapacité de fournir une information sur le mode initial d'obtention de la parcelle.

  - des différentes phases d'opérations de lotissement n'ont pas toujours été perçues de la même manière surtout en ce qui concerne le moment où une parcelle a été attribuée et celui où la zone a été considérée comme lotie.

  - des équipements de la parcelle et spécifiquement les équipements en eau où semble-t-il une confusion a eu lieu entre borne fontaine et forage.

Les données manquantes et imparfaites

Rappelons déjà qu'aucun refus de participer à l'enquête n'a eu lieu. Cependant des non-réponses ont été enregistrées. La plupart d'entre elles proviennent de l'incapacité de l'enquêté à répondre à la question parce qu'il n'a pas la connaissance du renseignement demandé (informations concernant la parcelle dont le chef de ménage résidant est un locataire ou un hébergé). Quelques unes de ces non-réponses sont dues à un oubli de l'enquêteur qui malgré les retours sur le terrain n'a pu obtenir une réponse (renseignements sur les bâtiments où un enquêteur n'a saisi ou cumulé qu'un bâtiment de la parcelle ce qui a échappé au contrôleur).

Mais les données manquantes peuvent se situer également sur un autre plan. Celui d'avoir omis d'inclure dans le questionnaire une question d'intérêt pour l'étude comme la date de lotissement ou la relation de parenté entre les chefs de ménage résidant sur une même parcelle. Ce dernier point rejoint les critiques de l'enquête.

Au dernier stade de l'enquête, nous avons pris réellement conscience des censures de l'information de la part des Burkinabès et surtout des Mossis, de notions de l'argent ou du temps qu'ils appréhendent différemment. Si ces constatations ont pu être observées lors de la préparation de l'enquête et prévenues au niveau du questionnaire, il n'en reste pas moins qu'il a fallu être très vigilant lors du déroulement de l'enquête. Il semble que parfois nous n'avons pas suffisamment insisté, auprès des enquêteurs et des superviseurs, sur la définition de certains concepts ou notions qui, *a priori*, ne paraissaient pas équivoques. Les phases de contrôle des questionnaires et de dépouillement ont permis de relever ces défauts et de les réajuster par une information résultant de combinaisons avec d'autres questions dont les réponses ont été cohérentes et satisfaisantes sur l'ensemble de l'enquête.

Mais il est certain que toutes les notions de situations un tant soit peu illégales (concernant entre autre le foncier ou les activités), les réserves propres aux Burkinabés par rapport à leur environnement ou de notions qu'ils appréhendent différemment (revenus, naissance et de décès) n'ont pu être mises à jour. Un des faits importants à noter est la stratégie foncière qui est inscrite sur le terrain mais non dite. Il a été difficile d'obtenir une information se distinguant d'un schéma imposé aussi bien par la politique urbaine que par un contexte environnant africain. Ces notions ont été cependant acquises par l'exploration du terrain, une connaissance et une pratique au préalable du milieu africain et interviennent largement dans l'interprétation des résultats.

Il ressort comme premier bilan que si une enquête demande énormément de précautions dans sa conception, lorsque celle-ci se déroule en milieu africain les difficultés spécifiques au terrain se multiplient: l'échantillonnage et l'existence de biais propres aux données d'enquêtes africaines. En effet, un terrain comme celui de Ouagadougou est amené à se modifier rapidement et ne possède pas de statistiques de référence. Aussi, si tous les ouvrages de sondages ou de méthodologies d'enquête insistent sur une rigueur dans la conception et la réalisation, nous restons malgré tout confrontés à des problèmes de méthode. On peut donc insister sur le fait qu'il existe un décalage entre la conceptualisation d'une enquête (phase théorique) et son déroulement (phase pratique), car il s'agit d'être rigoureux dans une situation qui ne favorise pas la rigueur.

Annexe 1.4 : Quelques principes d'analyse

Les variables

Création de variables

A partir des variables issues du questionnaire, s'est imposée la création d'autres variables plus adaptées aux objectifs de l'étude, obtenues par combinaison. Deux raisons motivent cette démarche:

- la première a été volontaire dans la conception même du questionnaire. D'une part, la création de nouvelles variables permet l'accès à des informations difficilement observables directement (la structure familiale). D'autre part, la conséquence des deux passages entraîne inévitablement une information traduisant l'évolution (taux de croissance, tissu urbain, investissement immobilier);

- la deuxième raison tient à la structure hiérarchique. Il était opportun de mettre en évidence certaines informations relatives aux unités de rang inférieur dans les unités hiérarchiquement supérieures (migration intra-urbaine, date et lieu d'arrivée à Ouagadougou, variables issues des unités "migrations" et intégrées au niveau des ménages...).

La modalité "non concerné"

Deux raisons expliquent la présence de cette modalité :

- L'existence des questions emboîtées et des deux passages d'enquête a suscité de nombreuses réponses "non concerné" qui, de par leur code, ont été différenciées des non-réponses.

  A l'exception des variables issues des caractéristiques de la parcelle et des variables identifiantes portant sur les ménages et individus, toutes les autres variables possèdent cette modalité.

- De par la procédure de la saisie et de son programme de mise à jour pour l'enquête à passages répétés, toutes les variables n'ont pas été introduites au second passage. Celles, spécifiques du premier passage, ont été affectées de la modalité "non concerné" au deuxième passage.

Cette modalité va conditionner le traitement statistique dans la mesure où il s'agira de travailler presque toujours sur des sous-populations.

L'analyse des groupes résidentiels

Cette analyse fait référence aux niveaux d'enquête. C'est la population "Ménage" qui a été choisie pour aborder le système résidentiel et ceci de trois manières :

- par la création de variables obtenues sur les unités individus (épouse, enfant, dépendant) et ramenées à l'unité Ménage. Ainsi, six groupes résidentiels sont définis suivant le statut matrimonial du chef de ménage et la présence ou non des dépendants:
  . groupe de résidence mononucléaire :
    - famille élémentaire simple,
    - famille élémentaire composée,
    - famille élémentaire simple élargie,
    - famille élémentaire composée élargie,
    - ménage de célibataire (marié ou non);
  . groupe de résidence polynucléaire.

Ces groupes peuvent être complétés suivant la présence ou non des enfants, des épouses sur la parcelle et le lien de parenté du dépendant avec le chef de ménage.

- par la répétition de l'information spécifique de la parcelle sur chaque ménage afin d'étudier, sur la population des ménages, les interactions entre la parcelle et le ménage.

- par la répétition également de l'information propre au Ménage sur l'unité Individu, l'analyse étant alors réalisée sur la population Individu.

L'analyse diachronique

Elle fait référence aux deux types de domaines d'étude suivant que l'on considère l'étude biographique (enquête rétrospective) ou l'étude des deux passages.

- L'étude biographique a été traitée par les niveaux d'enquête en considérant la relation "ménage-migration". On prend donc en considération la structure hiérarchique de l'information. L'enquête rétrospective ne concerne que les ménages enregistrés au premier passage. Nous nous sommes, entre autres, intéressés aux itinéraires migratoires hors de la ville et dans la ville. Pour cela, nous avons créé des variables sur l'unité Ménage à partir de l'unité migration.

  Par exemple, en combinant deux critères, celui de la naissance à Ouagadougou ou hors de Ouagadougou et celui de l'existence ou non d'une étape migratoire hors de la capitale, on aboutit à la définition de quatre types d'insertion dans la ville :
  - les natifs de Ouagadougou n'ayant jamais migré hors de la ville,
  - les non natifs de Ouagadougou n'ayant jamais migré hors de la ville,
  - les natifs de Ouagadougou ayant migré au moins une fois hors de la ville,
  - les non natifs de Ouagadougou ayant migré au moins une fois hors de la ville.

  De la même manière, nous saisissons la mobilité intra-urbaine suivant les itinéraires "centre-périphérie" et "périphérie-périphérie".

- l'étude des deux passages fait appel aux domaines d'étude définis par une variable. Les sous-populations obtenues à l'issue du deuxième passage font ressortir les ménages stables, les ménages arrivés et les ménages partis, auxquels ont été posées des questions spécifiques (raison d'arrivée ou de départ, lieu d'arrivée ou de départ pour les ménages migrants, changement de statut d'activité ou mobilité interne au ménage pour les ménages stables...). On a conduit une étude comparative dans le temps, à partir de deux tableaux qui ont une structure différente (variables différentes d'une passage à l'autre), ce qui a permis, d'une part d'étudier séparément ou conjointement les différentes sous-populations résultant des deux passages et, d'autre part de créer des variables décrivant l'évolution.

Annexe 1.5 : Le matériel informatique

Quatre impératifs sont exigés compte tenu des contraintes données par la conception même de la structure de l'enquête (passages répétés, hiérarchie du recueil des données) et orientent le choix du matériel informatique:

- **Un logiciel conçu pour micro-ordinateur** afin de pouvoir saisir et traiter immédiatement l'information sur le lieu de l'enquête.
- En fonction de ce premier impératif, un logiciel **le moins limité** possible dans le nombre d'enregistrements des variables, des individus et dans les capacités de composer la tabulation. Il s'agit de n'être contraint à aucun dimensionnement (environ 300 variables et 13000 individus, toutes unités confondues, doivent être enregistrés).
- Un logiciel permettant de saisir les informations aussi complexes soient-elles dans l'organisation du questionnaire qui peut se présenter sous la forme de différents niveaux d'enquête. Il s'agit d'**enregistrer les données** sans redondance des unités supérieures et **pour un nombre quelconque de niveaux.**
- Un logiciel capable de recevoir l'enregistrement de plusieurs passages d'enquête.

Ces conditions réunies sont rarement rencontrées et peu de logiciels les prennent toutes en considération. Celui qui a été choisi est un gestionnaire de données numériques hiérarchisées destiné au dépouillement d'enquêtes qui a permis de répondre à nos exigences. Conçu par J.VAUGELADE[1] et largement amélioré pour les besoins de notre enquête, il est écrit en BASIC sous MSDOS pour micro-ordinateur compatible IBM. Il présente:

- une description de la structure de l'enquête. Il s'agit de définir les types d'unité, leur niveau dans la hiérarchie, et de constituer les indications nécessaires pour décrire les variables qui sont affectées à ces types d'unités (libellé, nom abrégé, valeurs possibles de code).
- une saisie hiérarchique et une mise à jour. La saisie est effectuée de manière interactive avec contrôle et correction immédiate. Elle se particularise par une structure hiérarchique et qui offre deux avantages : une saisie simple compte tenu de la complexité du questionnaire pouvant faire intervenir différents niveaux d'enquête et l'occupation du minimum de place dans les fichiers d'enregistrement. Le programme de mise à jour a une double fonction : enregistrer plusieurs passages d'enquête et se déplacer dans le fichier.
- des contrôles de qualité. La saisie est simultanément accompagnée d'un pré-traitement. Un premier contrôle nous assure que chaque réponse est possible car elle appartient à une catégorie de codes précisés. Un deuxième contrôle, dit croisé, permet de vérifier la pertinence des valeurs de la variable en fonction des valeurs d'une ou de plusieurs autres variables. Il aboutit à un choix de sept types de contrôles fondés sur des tests logiques.
- la création de variables.
- la création de tableaux croisés. Ce module permet de composer des tableaux de contingence ou de fréquence à partir d'une ou de plusieurs variables et de travailler sur une population pondérée ou sur des sous-populations sélectionnées par le principe de filtres.

Il est certain que d'autres points sont moins bien développés notamment une bonne présentation des résultats et l'absence de graphiques. Mais si ceux-ci ne doivent pas non plus être négligés, ils restent néanmoins secondaires par rapport à une obligation de conception et d'enregistrement qui, dans notre cas, est la saisie hiérarchique.

[1] Démographe et informaticien au centre Orstom de Ouagadougou.

Annexe 2

COMPLEMENTS STATISTIQUES POUR L'ANALYSE DU SYSTEME D'ECHELLES

Annexe 2.1 : Les tableaux de "Burt"

Notations de base

Soient les populations d'étude suivantes :

- A, l'ensemble des unités statistiques du type A . $A = \cup \{ a \mid a \in A \}$; Card (A) = CA.
- B, l'ensemble des unités statistiques du type B . $B = \cup \{ b \mid b \in B \}$; Card (B) = CB.
- C, l'ensemble des unités statistiques du type C . $C = \cup \{ c \mid c \in C \}$; Card (C) = CC.
- D, l'ensemble des unités statistiques du type D . $D = \cup \{ d \mid d \in D \}$; Card (D) = CD.

Chaque population correspond à un niveau dans l'arbre hiérarchique, A, B, C et D.

On note les ensembles de variables et de modalités:

- Q_A, l'ensemble des variables caractérisant les unités statistiques du type A.
  $Card(Q_A) = CQ_A$.
  On note sur le même principe, Q_B, Q_C et Q_D.
- J_A, l'ensemble des modalités associées aux variables de Q_A et l'on a de la même manière J_B, J_C et J_D.
  On pose : $Q = Q_A \cup Q_B \cup Q_C$ et $J = J_A \cup J_B \cup J_C$

Construction des tableaux de Burt ou assimilés en tant que tels

On définit à partir des tableaux HC, HB, HA, une série de tableaux de correspondances symétriques que l'on appelle, par abus de langage, tableaux de Burt BC(JJ), BB(JJ) et BA(JJ). Leurs termes génériques et leurs marges sont donnés par :

$$\left.\begin{aligned} BC_{jj'} &= \sum_c \frac{HC_{cj}\, HC_{cj'}}{HC_c} \quad \text{et} \quad BC_j = HC_j * CQ \\ BB_{jj'} &= \sum_b \frac{HB_{bj}\, HB_{bj'}}{HB_b} \quad \text{et} \quad BB_j = HC_j * CQ \\ BA_{jj'} &= \sum_a \frac{HA_{aj}\, HA_{aj'}}{HA_a} \quad \text{et} \quad BA_j = HC_j * CQ \end{aligned}\right] \qquad [2]$$

On sait alors que l'analyse des correspondances du tableau BC (resp. BB et BA) est équivalente à celle du tableau HC (resp. HB et HA)[1].

[1] [BENZECRI, 1973] T IIA p243.

Comparaison des blocs de groupes de variables issus des tableaux de Burt

Rappelons que du fait de l'emboîtement des unités statistiques et donc de la réaffectation des informations des niveaux supérieurs sur les niveaux inférieurs, nous avons :

$$\forall j \in J_A \text{ et } \forall c \in a : \quad HC_{cj} = \frac{HA_{aj}}{p_a}$$

$$\forall j \in J_A \cup J_B \text{ et } \forall c \in b : \quad HC_{cj} = \frac{HB_{bj}}{p_b}$$

Par conséquent on démontre[1] que :

$$\forall j \in J_A \text{ et } \forall j' \in J : \quad BC_{jj'} = BB_{jj'} = BA_{jj'}$$

$$\text{et } \forall j \in J_B \text{ et } \forall j' \in J : \quad BC_{jj'} = BB_{jj'}$$

Ainsi, les analyses à différents niveaux du bloc J_A*J sont équivalentes et il en est de même du bloc J_B*J pris aux niveaux intermédiaire et inférieur, ce qui se présente comme :

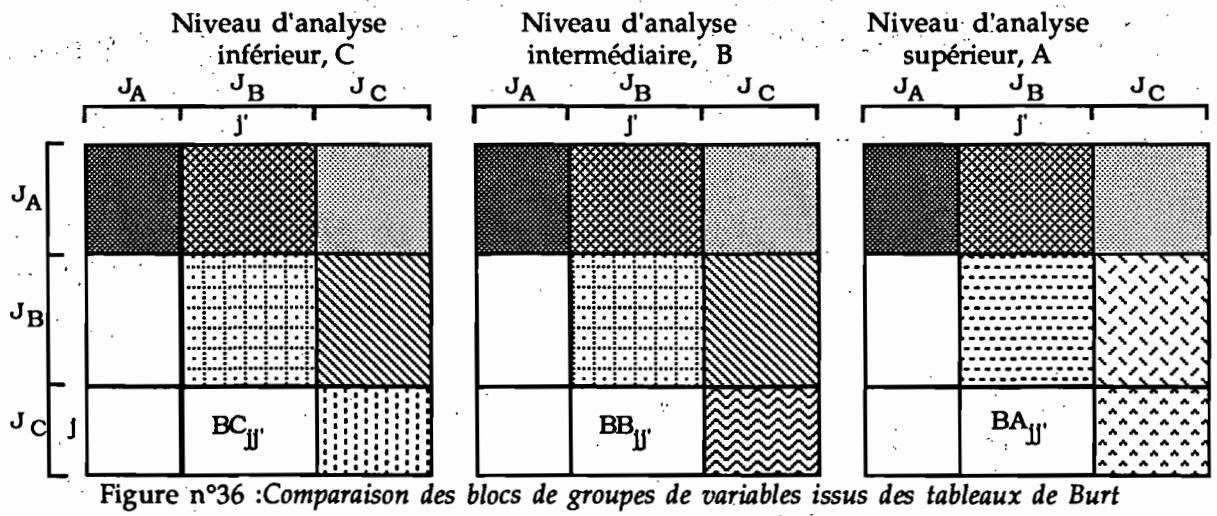


Figure n°36 :*Comparaison des blocs de groupes de variables issus des tableaux de Burt*

Les marges de ces tableaux sont semblables et sont données par : $\forall j \in J : BC_j = BB_j = BA_j$

[1][PIRON, 1990]

Interprétation des blocs de groupes de variables

Nous allons, maintenant, examiner les termes des différents blocs pris sur les trois niveaux:

- Les blocs J_A*J où nous avons quelque soit le niveau: $BC_{jj'} = BB_{jj'} = BA_{jj'}$
  . pour J_A*J_C, ce terme exprime le nombre d'individus c ayant la modalité j' de J_C et contenus dans les unités a qui possèdent la modalité j de J_A.
  . pour J_A*J_B, ce terme exprime le nombre d'individus c contenus dans les unités b de B qui possèdent la modalité j' de J_B et qui sont incluses dans les unités a de A lesquelles possèdent la modalité j de J_A.
  . pour J_A*J_A, ce terme exprime le nombre d'individus c contenus dans les unités a lesquelles possèdent à la fois les modalités j et j' de J_A.

- Les blocs J_B*J_B où nous avons :
  . sur les niveaux C et B, $BC_{jj'} = BB_{jj'}$. C'est le nombre d'individus c contenus dans les unités b lesquelles possèdent à la fois les modalités j et j' de J_B.
  . sur le niveau A, $BA_{jj'}$ est le nombre d'individus c contenus dans les unités b qui possèdent à la fois les modalités j et j' de J_B et qui sont incluses dans une même unité a.

- Les blocs J_B*J_C où nous avons :
  . sur les niveaux B et C, $BC_{jj'} = BB_{jj'}$. C'est le nombre d'individus c ayant la modalité j' de J_C et qui sont contenus dans les unités b lesquelles possèdent la modalité j de J_B.
  . sur le niveau A, $BA_{jj'}$ qui est le nombre d'individus c ayant la modalité j' de J_C et contenus dans les unités b qui possèdent la modalité j de J_B et qui sont incluses dans une même unité a.

- les blocs J_C*J_C :
  . sur le niveau C, nous avons $BC_{jj'}$ qui désigne le nombre d'individus c ayant les modalités j et j' de J_C.
  . sur le niveau B, $BB_{jj'}$ est le nombre d'individus qui possèdent les modalités j et j' de J_C et qui sont incluses dans une même unité b.
  . sur le niveau A, $BA_{jj'}$ est le nombre d'individus ayant les modalités j et j' de J_C et qui sont contenus dans une même unité a.

Annexe 2.2 : La décomposition de l'inertie pour une série de partitions emboîtées

Les différents types de nuages de points emboîtés

Les différents niveaux définissent des partitions sur les populations de niveaux inférieurs et nous allons décomposer le nuage de points de l'ensemble des individus de C muni de deux partitions emboîtées, celle engendrée par A et celle induite par B.

Partant de la structure hiérarchique de base le tableau disjonctif complet pondéré par p_C, HC(C,J), on définit sur l'ensemble C*J une mesure par la loi de probabilité :

$$F_{CJ} = \{F_{cj} \mid c \in C ; j \in J\} ; \; F_{cj} = \frac{HC_{cj}}{HC} \text{ où } HC = \sum_c \sum_j HC_{cj} = CJ*CC$$

Les lois marginales et conditionnelles s'expriment par :

$$F_C = \{F_c = \frac{HC_c}{HC} \mid c \in C\} \text{ et } F_J = \{F_j = \frac{HC_j}{HC} \mid j \in J\} ;$$

$$F_J^C = \{F_j^c = \frac{F_{cj}}{F_c} \mid c \in C ; j \in J\} \text{ et } F_C^J = \{F_c^j = \frac{F_{cj}}{F_j} \mid c \in C ; j \in J\}$$

On considère également la mesure sur A*J définie à partir de HA(A,J) :

$$F_{AJ} = \{F_{aj} \mid a \in A ; j \in J\} ; \; F_{aj} = \frac{HA_{aj}}{HA} \text{ et } F_A = \{F_a = \frac{HA_a}{HA} \mid a \in A\}$$

et la mesure sur B*J définie à partir de HB(B,J) :

$$F_{BJ} = \{F_{bj} \mid b \in B ; j \in J\} ; \; F_{bj} = \frac{HB_{bj}}{HB} \text{ et } F_B = \{F_b = \frac{HB_b}{HB} \mid b \in B\}$$

Rappelons que HA = HB = HC .

Soit N(C), le nuage de points de R_J espace des modalités muni de la distance du $\chi 2$ de centre F_J : $N(C) = \{F_J^c ; F_c \mid c \in C\} ; \sum_c F_c = 1 ; F_J$ est le centre de gravité du nuage N(C). On définit alors une série de nuages emboîtés issus des partitions de A et de B sur C. Considérons d'abord les simples partitions sur C:

- $C = \cup \{C_a \mid a \in A\}$.

  le nuage des centres de gravités des classes sur C pour une partition induite par A : $N(A) = \{F_J^a ; F_a \mid a \in A\}$ où $F_a = \frac{p_a}{CC}$; F_J est le centre de gravité du nuage N(A).

  le nuage des points d'une classe a de A, C_a : $\forall a \in A : N(C_a) = \{F_J^c ; F_{(ca)} \mid c \in a\}$ où $F_{(ca)} = \frac{p_c}{p_a}$;

  F_J^a est le centre de gravité du nuage $N(C_a)$.

- $C = \cup \{C_b \mid b \in B\}$.

  le nuage des centres de gravités des classes associées à la partition engendrée par B sur C: $N(B) = \{F_J^b ; F_b \mid b \in B\}$ où $F_b = \frac{p_b}{CC}$; F_J est le centre de gravité du nuage N(B).

le nuage des points d'une classe b de B, C_b : $\forall b \in B : N(C_b) = \{F_J^c ; F_{(cb)} \mid c \in b\}$ où $F_{(cb)} = \frac{p_c}{p_b}$;

F_J^b est le centre de gravité du nuage $N(C_b)$.

Regardons maintenant la série de nuages emboîtés issus de la double partition :

$$C = \cup \{\cup \{C_b \mid b \in B_a\} \mid a \in A\}$$

Considérons pour cela la partition : $C_a = \cup \{C_b \mid b \in B_a\}$.

- soit $N(B_a)$, le nuage des centres de gravité des classes associées à la partition engendrée par B_a sur C_a :

  $\forall a \in A : N(B_a) = \{F_J^b ; F_{(ba)} \mid b \in B_a\}$ où $F_{(ba)} = \frac{p_b}{p_a}$; F_J^a est le centre de gravité du nuage $N(B_a)$.

- soit $N(C_b)$, le nuage des points d'une classe de b de a :

  $\forall a \in A$ et $\forall b \in a : N(C_b) = \{F_J^c ; F_{(cb)} \mid c \in C_b\}$ où $F_{(cb)} = \frac{1}{p_b}$; F_J^b est le centre de gravité du nuage $N(C_b)$.

Par dualité, on a N(J), le nuage de points de R_C espace des individus muni de la distance du $\chi 2$ de centre F_J : $N(J) = \{F_C^j ; F_j \mid j \in J\}$; $\sum_j F_j = 1$; F_C est le centre de gravité du nuage N(J) dans R_C. N(J) peut être également considéré d'une part dans R_B muni de la distance du $\chi 2$ de centre F_B : $N(J) = \{F_B^j ; F_j \mid j \in J\}$; F_B est le centre de gravité du nuage N(J) dans R_B et d'autre part dans R_A muni de la distance du $\chi 2$ de centre F_P : $N(J) = \{F_A^j ; F_j \mid j \in J\}$; F_A est le centre de gravité du nuage N(J) dans R_A.

Schéma de décomposition pour une simple partition

Pour les simples partitions, l'inertie du nuage N(C) se décompose suivant :

- la partition sur C engendrée par A : $C = \cup \{C_a \mid a \in A\}$.

$$In\,(N(C)) = \sum_a \frac{F_a}{F_J} (F_J^a - F_J)^2 + \sum_a \sum_{c \in a} \frac{F_{(ca)}}{F_J} (F_J^c - F_J^a)^2$$

d'où $$In\,(N(C)) = In\,(N(A)) + \sum_a In\,(N(C_a))$$

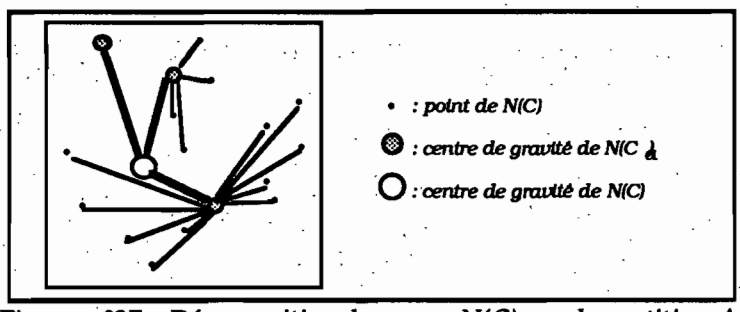


Figure n°37 : *Décomposition du nuage N(C) par la partition A*

- la partition sur C engendrée par B : $C = \cup \{ C_b \mid b \in B \}$:

$$In\ (N(C)) = \sum_b \frac{F_b}{F_J} (F_J^b - F_J)^2 + \sum_b \sum_{c \in b} \frac{F_{(cb)}}{F_J} (F_J^c - F_J^b)^2$$

d'où $$In\ (N(C)) = In\ (N(B)) + \sum_b In\ (N(C_b))$$

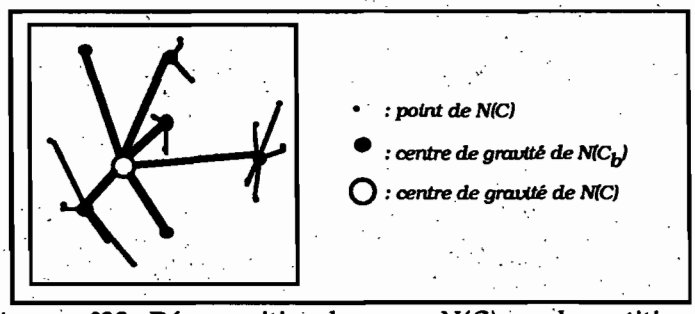


Figure n°38 :*Décomposition du nuage N(C) par la partition B*

- la partition sur B engendrée par A : $B = \cup \{ B_a \mid b \in a \}$. B est l'ensemble des centres de gravité des classes associée à la partition B sur C . On a :

$$In\ (N(B)) = \sum_a \frac{F_a}{F_J} (F_J^a - F_J)^2 + \sum_a \sum_{b \in a} \frac{F_{(ba)}}{F_J} (F_J^b - F_J^a)^2$$

d'où $$In\ (N(B)) = In\ (N(A)) + \sum_a In\ (N(B_a))$$

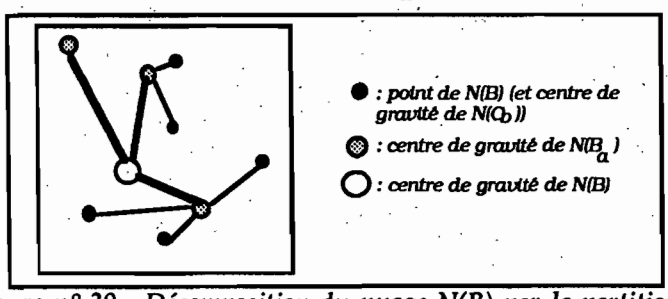


Figure n° 39 : Décomposition du nuage N(B) par la partition A

Schéma de décomposition pour une double partition

On considère la double partition sur C : $C = \cup \{ \cup \{C_b \mid b \in B_a\} \mid a \in A \}$. Regardons d'a@bord comment se décompose l'inertie sur le sous-nuage des points d'une classe a de A, $N(C_a)$ où :

$$C_a = \cup \{ C_b \mid b \in B_a \}$$

On a : $$In\,(N(C_a)) = \sum_{b \in a} \frac{F_{(ba)}}{F_J} (F_J^b - F_J^a)^2 + \sum_{b \in a} \sum_{c \in b} \frac{F_{(cb)}}{F_J} (F_J^c - F_J^b)^2$$

d'où $$In\,(N(C_a)) = In\,(N(B_a)) + \sum_{B_a} In\,(N(C_b))$$

Ainsi, on a d'une manière équivalente en considérant les simples partitions sur C :

$$In\,(N(C)) = In\,(N(A)) + \sum_{a} In\,(N(C_a))$$

et $$In\,(N(C)) = In\,(N(B)) + \sum_{b} In\,(N(C_b))$$

et on en déduit :

$$In\,(N(C)) = In\,(N(A)) + \sum_{a} In\,(N(B_a)) + \sum_{a} \sum_{b \in a} In\,(N(C_b))$$

c'est-à-dire

$$In\,(N(C) = \sum_{a} \frac{F_a}{F_J} (F_J^a - F_J)^2 + \sum_{a} \sum_{b \in a} \frac{F_{(ba)}}{F_J} (F_J^b - F_J^a)^2$$

$$+ \sum_{a} \sum_{b \in a} \sum_{c \in b} \frac{F_{(cb)}}{F_J} (F_J^c - F_J^b)^2$$

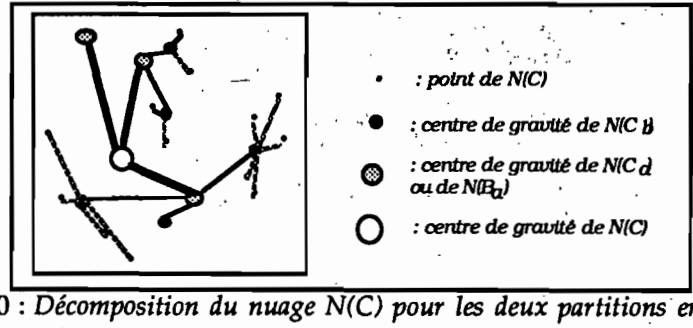


Figure n° 40 : *Décomposition du nuage N(C) pour les deux partitions emboîtées sur C*

Annexe 2.3 : Expression de la décomposition de l'inertie d'après le tableau de Burt

La décomposition de l'inertie peut s'envisager d'après les termes généraux des tableaux de Burt, BC, BB, BA associés aux tableaux de correspondances HC, HB, HA.

Soulignons pour cela qu'un terme général du tableau de Burt peut s'exprimer en terme de covariance de la façon suivante :

$$BC_{jj'} = \sum_c \frac{F_{cj}\, F_{cj'}}{F_c} = \sum_c F_c\, F_j^c\, F_{j'}^c = F_j F_{j'} + \sum_c F_c\, (F_j^c - F_j)\, (F_{j'}^c - F_{j'})$$

et l'on a de la même manière :

$$BB_{jj'} = F_j F_{j'} + \sum_b F_b\, (F_j^b - F_j)\, (F_{j'}^b - F_{j'})$$

$$BA_{jj'} = F_j F_{j'} + \sum_a F_a\, (F_j^a - F_j)\, (F_{j'}^a - F_{j'})$$

Pour une simple partition

Pour une simple partition engendrée sur C par A par exemple, on obtient une formule de décomposition du terme $BC_{jj'}$ donnée en fonction de la décomposition de l'inertie. En effet,

$$BC_{jj'} = F_j F_{j'} + \sum_a \sum_{c \in a} F_c\, (F_j^c - F_j)\, (F_{j'}^c - F_{j'})$$

$$= F_j F_{j'} + \sum_a \sum_{c \in a} F_c\, (F_j^c - F_j^a + F_j^a - F_j)\, (F_{j'}^c - F_{j'}^a + F_{j'}^a - F_{j'})$$

$$= F_j F_{j'} + \sum_a F_a\, (F_j^a - F_j)\, (F_{j'}^a - F_{j'}) + \sum_a \sum_{c \in a} F_c\, (F_j^c - F_j^a)\, (F_{j'}^c - F_{j'}^a)$$

En posant : $$BCI_{jj'} = \sum_a \sum_{c \in a} F_c\, (F_j^c - F_j^a)\, (F_{j'}^c - F_{j'}^a)$$

on obtient : $$BC_{jj'} = BA_{jj'} + BCI_{jj'}$$

autrement dit : $$BC_{jj'} - BC_j\, BC_{j'} = (BC_{jj'} - BA_{jj'}) + (BA_{jj'} - BC_j\, BC_{j'})$$

Remarque :

- On peut faire référence à J.J. DAUDIN[1] qui exprime, dans le cas de tableaux ternaires, cette dernière relation de la façon suivante :
  "dépendance marginale = dépendance détachée de A + dépendance attachée à A"

[1] [DAUDIN, 1981] p17.

Pour une double partition

Pour deux partitions emboîtées A et B agissant sur l'ensemble C on a la décomposition suivante:

$$BC_{jj'} = F_j F_{j'} + \sum_a \sum_{b \in a} \sum_{c \in a} F_c \, (F^c_j - F_j) \, (F^c_{j'} - F_{j'})$$

$$= F_j F_{j'} + \sum_a F_a \, (F^a_j - F_j) \, (F^a_{j'} - F_{j'}) + \sum_a \sum_{b \in a} F_b \, (F^b_j - F^a_j) \, (F^b_{j'} - F^a_{j'}) + \sum_b \sum_{c \in b} F_c \, (F^c_j - F^b_j) \, (F^c_{j'} - F^b_{j'})$$

en posant : $$BBI_{jj'} = \sum_a \sum_{b \in a} F_b \, (F^b_j - F^a_j) \, (F^b_{j'} - F^a_{j'})$$

et $$BCI'_{jj'} = \sum_b \sum_{c \in b} F_c \, (F^c_j - F^b_j) \, (F^c_{j'} - F^b_{j'})$$

on obtient : $$BC_{jj'} = BA_{jj'} + BBI_{jj'} + BCI'_{jj'}$$

où $$BCI'_{jj'} = BC_{jj'} - BB_{jj'}$$

et $$BBI_{jj'} = BB_{jj'} - BA_{jj'}$$

soit encore :

$$BC_{jj'} - BC_j \, BC_{j'} = (BC_{jj'} - BB_{jj'}) + (BB_{jj'} - BA_{jj'}) + (BA_{jj'} - BC_j \, BC_{j'})$$

Annexe 2.4 : Rappel des principaux travaux sur l'analyse intra-classes

Nous rappelons ici les principales études effectuées sur l'analyse intra-classes qui diffèrent selon la nature du tableau des données :

- La plupart des études faites sur la décomposition de l'inertie se basent sur une décomposition simple, c'est-à-dire en deux éléments, d'un **tableau ternaire**. J.P. BENZECRI propose un tableau correspondant à l'inertie intra-classes et de terme général:

  $$"kr(i,j,t) = k(i,j,t) - \frac{k(i,j)\,k(i,t)}{k(i)} + \frac{k(j)}{k} k(i,t) \ "$$

  Effectuer une analyse factorielle des correspondances sur un tel tableau revient à effectuer l'analyse simultanée des sous-nuages des profils-lignes (sur lequel a été opéré la partition) recentrés à l'origine. En effet, "*le nuage associé à l'inertie intra-classes est construit en disposant dans* R^J *autour du centre de gravité du nuage total les vecteurs différences entre chaque élément de* I_t *et le centre de sa classe*"[1].

- Pour traiter la décomposition de l'inertie sur un **tableau disjonctif complet**, B.ESCOFIER propose l'analyse des correspondances multiples conditionnelle[2] c'est-à-dire une analyse intra-classes effectuée sur un tableau disjonctif complet. L'objectif est "*d'éliminer l'influence de la variable T* (celle qui constitue une partition) *extérieure aux données traitées pour une analyse donnée*". Pour éliminer d'une part dans le nuage N(I) la dispersion due à la variable T et d'autre part dans le nuage N(J) la part provenant de la liaison entre les questions et la variable T, il faut analyser le nuage des individus recentrés par classe dans R^J et le nuage de modalités projeté sur l'orthogonal du sous-espace, E_T , engendré par les indicatrices des modalités de T. Ces deux transformations sont duales l'une de l'autre. Ceci revient alors à réaliser une analyse des correspondances sur le tableau de terme général :

  $$"k^*_{ij} = k_{ij} - \frac{b_{jt}}{k_t} + \frac{k_j}{n} \quad si \quad i \in I_t " \qquad \text{et où } b_{jt} = \sum_{I_t} k_{ij}$$

- L'analyse des correspondances internes d'un tableau partitionné[3] rend compte de l'étude de deux partitions agissant à la fois sur I et sur J. Il s'agit d'un **tableau de correspondances** T dont le tableau de fréquences relatives est noté P et où l'on considère :
  - d'une part une partition des lignes sous la forme d'une variable qualitative à L modalités qui engendre un tableau disjonctif complet X comportant I lignes et L colonnes
  - d'autre part une partition des colonnes sous la forme d'une variable qualitative à K modalités qui engendre un tableau disjonctif complet Y comportant J lignes et K colonnes.

  Tout comme pour l'analyse des correspondances multiples conditionnelle, on projette d'une part le nuage N(I) sur l'orthogonal du sous-espace engendré par les colonnes de Y dans R^J et d'autre part le nuage N(J) sur l'orthogonal sur du sous-espace engendré par les colonnes de X dans R^I. Ainsi "*l'analyse des correspondances internes est équivalente à l'analyse factorielle des correspondances du tableau R de terme général :* $r_{ij} = p^\circ_{ij} + p_{i.}\ p_{.j}$ "[4]

  où $$p^\circ_{ij} = p_{ij} - p_{lj}\frac{p_i}{p_l} - p_{ik}\frac{p_j}{p_k} + p_{lk}\frac{p_i}{p_l}\frac{p_j}{p_k} \quad \text{et} \quad p_l = \sum_{i \in l} p_i \ ; \ p_k = \sum_{j \in k} p_j$$

---

[1] [BENZECRI, 1983] p357.

[2] [ESCOFIER, 1987b]

[3] [CAZES *et al*, 1988]

[4] [CAZES *et al*, 1988] p 46.

Annexe 2.5 : L'analyse inter-classes

Deux analyses inter-classes sur le nuage N(C) sont envisagées suivant que l'on considère la partition sur C engendrée par A ou par B. Nous rappelons les formules suivant le tableau des centres de gravité, le tableau de correspondances et le tableau de Burt associé à ces deux premiers tableaux:

- d'une part, l'inertie inter-classes pour la partition engendrée par A sur C c'est-à-dire l'inertie du nuage des centres de gravité des classes de A :

$$In(N(A)) = \sum_a \frac{F_a}{F_J} (F_J^a - F_J)^2)$$

Le tableau associé à l'inertie inter-classes peut être envisagé d'une manière équivalente par $HC^{\circ\circ}(C,J)$, HA(A,J) ou BA(J,J) de termes généraux respectifs :

$\forall c \in a$:

$$HC^{\circ\circ}_{cj} = \frac{HA_{aj}}{p_a}$$

$$HA_{aj} = \sum_{c \in a} HC_{cj}$$

$$BA_{jj'} = \sum_a \frac{HA_{aj}\, HA_{aj'}}{HA_a}$$

- d'autre part, l'inertie inter-classes pour la partition induite par B sur C :

$$In(N(B)) = In\,(N(A)) + \sum_a In\,(N(B_a))$$

$$In(N(B)) = \sum_b \frac{F_b}{F_J} (F_J^b - F_J)^2)$$

Le tableau associé à l'inertie inter-classes peut être envisagé d'une manière équivalente par HC°_{cj}, HB(B,J) ou BB(J,J) de termes généraux :

$\forall c \in b$:

$$HC^{\circ}_{cj} = \frac{HB_{bj}}{p_b} ;$$

$$HB_{bj} = \sum_{c \in b} HC_{cj} ;$$

$$BB_{jj'} = \sum_b \frac{HB_{bj}\, HB_{bj'}}{HB_b} ;$$

Remarque :

- L'expression du tableau associé à l'inertie inter-classes donnée par le tableau de Burt exprime l'indépendance entre j et j' conditionnellement à la partition A ou B. On retrouve alors l'analogie avec ce que J.J. DAUDIN[1] appelle *"les dépendances attachées à la partition"* .

[1] [DAUDIN, 1981] p21.

Annexe 2.6 : L'analyse intra-classes

Sur le même principe que pour l'analyse inter-classes, nous considérons les deux analyses intra-classes issues respectivement de la partition sur C engendrée par A et celle induite par B. La construction des tableaux associés à ces deux inerties est fondée sur l'analyse des correspondances multiples conditionnelle que propose B.ESCOFIER. Nous avons :

- d'une part, l'inertie intra-classes pour la partition engendrée par A. C'est donc la somme sur A des inerties des nuages $N(C_a)$) :

$$N^{**}(C) = \sum_a In\,(N(C_a)) = \sum_a [In\,(N(B_a)) + \sum_{b \in a} In\,(N(C_b))]$$

$$N^{**}(C) = \sum_a \sum_{c \in a} \frac{F_{(ca)}}{F_J} (F_J^c - F_J^a)^2)$$

Le tableau associé à cette inertie intra-classes, noté HC**(C,J), a pour terme général :

$$\forall c \in a : HC^{**}_{cj} = HC_{cj} - \frac{HA_{aj}\,HC_c}{HA_a} + \frac{HC_j\,HC_c}{HC} = HC_{cj} - \frac{p_c}{p_a} HA_{aj} + \frac{p_c}{CC} HC_j$$

d'où pour $p_c = 1$ $\qquad HC^{**}_{cj} = HC_{cj} - HC^{\circ\circ}_{cj} + \frac{HC_j}{CC}$

où HC_{cj} est le terme général du tableau de données HC(C,J), $HC^{\circ\circ}_{cj}$ est le terme général du tableau associé à l'inertie inter-classes HC°°(C,J) c'est-à-dire du tableau modèle et $\frac{HC_j}{CC}$ est le produit des marges. Les marges de H**(C,J) sont semblables à celles de HC(C,J) et de HC°°(C,J).

Les lois conditionnelles sont définies par:

$\forall c \in a$: $\qquad F^{**c}_J = F^c_J - F^{\circ\circ c}_J + F_J = F^c_J - F^a_J + F_J$

et $\qquad F^{**j}_C = F^j_C - F^{\circ\circ j}_C + F_C = F^j_C - \frac{F^j_A}{p_a} + F_C$

L'analyse de ce tableau étudie la dispersion des éléments de C_a à l'intérieur des classes de A.

- d'autre part, l'inertie intra-classes pour la partition engendrée par B qui est également la somme sur A des inerties intra-classes des nuages $N(C_a)$) :

$$N^*(C) = \sum_b In\,(N(C_b)) = \sum_a \sum_{b \in a} In\,(N(C_b))$$

$$N^*(C) = \sum_b \sum_{c \in b} \frac{F_{(cb)}}{F_J} (F_J^c - F_J^b)^2$$

On considère, de la même manière, le tableau associé à cette inertie intra-classes, noté HC*(C,J) de terme général et de marges :

$$\forall c \in b : HC^*_{cj} = HC_{cj} - \frac{HB_{bj}\,HC_c}{HB_b} + \frac{HC_j\,HC_c}{HC} = HC_{cj} - \frac{p_c}{p_b} HB_{bj} + \frac{p_c}{CC} HC_j$$

d'où pour $p_c = 1$ $\qquad HC^*_{cj} = HC_{cj} - HC^\circ_{cj} + \frac{HC_j}{CC}$

Construction du tableau de Burt associé à l'inertie intra-classes

On démontre que l'on peut construire les tableaux de Burt BC**(J,J) et BC*(J,J) associés respectivement aux tableaux hiérarchisés HC**(C,J) et HC*(C,J). En effet, on a pour la partition sur C engendrée par A :

$$\forall c \in a : \quad BC^{**}_{jj'} = \sum_c \frac{HC^{**}_{cj}\, HC^{**}_{cj'}}{HC^{**}_c} = \sum_a \sum_{c \in a} \frac{HC^{**}_{cj}\, HC^{**}_{cj'}}{HC^{**}_c}$$

où $\qquad HC^{**}_{cj} = HC_{cj} - \frac{HA_{aj}\, HC_c}{HA_a} + \frac{HC_j\, HC_c}{HC}$

on vérifie alors que : $\qquad BC^{**}_{jj'} = BC_{jj'} - BA_{jj'} + \frac{BC_j BC_{j'}}{CC}$

on a de la même manière, pour la partition sur C engendrée par B :

$$BC^*_{jj'} = BC_{jj'} - BB_{jj'} + \frac{BC_j BC_{j'}}{CC}$$

Remarques :

- On considère la partition sur C engendrée par A. $BC_{jj'}$ et $BA_{jj'}$ ont les mêmes marges et BA(J,J) exprime l'inertie inter-classes qui correspond au tableau modèle. *"Pour analyser simultanément les liaisons binaires entre les variables en éliminant l'influence de A"*[1], on réalise une analyse des correspondances sur le tableau BC**(J,J) de terme général :

  $$BC^{**}_{jj'} = BC_{jj'} - BA_{jj'} + \frac{BC_j BC_{j'}}{CC}$$

  En considérant la partition sur C engendrée par B on a de la même manière :

  $$BC^*_{jj'} = BC_{jj'} - BB_{jj'} + \frac{BC_j BC_{j'}}{CC}$$

  On retrouve ainsi l'expression de la décomposition de l'inertie (pour la partition engendrée par A) d'après un tableau de Burt où :

  $$BCI_{jj'} = BC^{**}_{jj'} - \frac{BC_j BC_{j'}}{CC} = BC_{jj'} - BA_{jj'}$$

  et ce que J.J. DAUDIN appelle *"la dépendance détachée de A"* [2].

- A partir de la construction des tableaux de Burt BC* et BC**, nous pouvons envisager l'analyse des sous-tableaux de Burt croisant les groupes de variables J_A, J_B, J_C . On rappelle que les marges sont proportionnelles à celles des tableaux de Burt correspondants.

---

[1] [ESCOFIER, 1987b]

[2] [DAUDIN, 1981] p22.

Annexe 2.7 : L'analyse de "l'inertie intermédiaire"

La décomposition de l'inertie sur trois niveaux a abouti à la considération de trois formes d'inertie. Nous venons de prendre en compte deux d'entre elles : l'inertie inter-classes pour la partition engendrée par A sur C et l'inertie intra-classes pour la partition induite par B sur C. Il s'agit maintenant de donner une signification à l'inertie intermédiaire qui découle de la série de partitions emboîtées et qui vaut :

$$\sum_a \text{In}\,(N(B_a)) = \sum_j \sum_a \sum_{b \in a} \frac{F_{(ba)}}{F_j} \left(F_j^b - F_j^a \right)^2$$

et où $\forall c \in b, \forall b \in a$: $F_j^b = F^{\circ c}_j$; $F_j^a = F^{\circ\circ c}_j$ et $F_{(ba)} = \dfrac{F^{\circ}_c}{F^{\circ\circ}_c} = \dfrac{p_b}{p_a}$

Construction du tableau associé

Il s'agit de l'inertie intra-classes pour la partition engendrée par A des centres de gravité des classes b associées à la partition B sur C. Mais c'est aussi la somme sur A des inerties inter-classes du nuage $N(C_a)$ partitionné par B_a. Par conséquent, l'analyse du tableau doit réaliser au préalable une analyse inter-classes du nuage $N(C_a)$ partitionné par B_a.

Dans le nuage N(C) des points de R_J, chaque élément c de C_b est donc assimilé au profil moyen de sa classe b, lequel est analysé par rapport au profil moyen du nuage $N(B_a)$. Notons que si les réponses aux questions varient beaucoup d'une classe a à une autre, ces sous-nuages sont séparés les uns des autres. On étudie ainsi l'effet de la partition la plus fine B sur les individus puis on élimine dans les sous-nuages $N(B_a)$ la dispersion due à la partition la moins fine A en les recentrant à leur origine. Dans le nuage N(J) des points de R_C , on élimine la part provenant de la liaison entre les variables Q et la variable de partition q_A . Cette part apparait dans la projection de N(J) sur l'orthogonal du sous-espace engendré par les indicatrices des modalités de q_A .

Pour cela, on crée d'abord le tableau associé à l'analyse inter-classes pour la partition induite par B sur C c'est-à-dire le tableau HC°(C,J) de terme général :

$\forall c \in b$: $$HC^{\circ}_{cj} = \sum_{c \in b} \frac{HC_{cj}}{p_b} = \frac{HB_{bj}}{p_b}$$

dont l'analyse est équivalente à celle réalisée sur le tableau HB(B,J).

On réalise ensuite l'analyse intra-classes du nuage $N(B_a)$ pour la partition engendrée par A c'est-à-dire l'analyse intra-classes du tableau HC°(C,J) (ou HB(B,J)). Dans ce cas, le tableau modèle doit traduire la dispersion inter-classes pour la partition A sur B et qui traduit également la dispersion inter-classes pour la partition A sur C. Il s'agit par conséquent du tableau HC°°(C,J) (ou HA(A,J)) de terme général :

$\forall c \in a$: $$HC^{\circ\circ}_{cj} = \sum_{c \in a} \frac{HC_{cj}}{p_a} = \frac{HA_{aj}}{p_a}$$

Notons $HC^{\sim}(C,J)$ le tableau associé à l'inertie intermédiaire c'est-à-dire à l'inertie intra-classes des centres de gravité des classes b pour la partition engendrée par A. Son terme général s'exprime par :

$$\forall c \in b \text{ et } \forall b \in a : \quad HC^{\sim}_{cj} = HC^{\circ}_{cj} - HC^{\circ\circ}_{cj} + \frac{HC_j}{CC} = \frac{HB_{bj}}{p_b} - \frac{HA_{aj}}{p_a} + \frac{HC_j}{CC}$$

HC°_{cj} est le terme général du tableau de correspondances $HC^{\circ}(C,J)$; $HC^{\circ\circ}_{cj}$ est le terme général du tableau modèle $HC^{\circ\circ}(C,J)$; $\frac{HC_j}{CC}$ est le produit des marges. Les marges de $HC^{\sim}(C,J)$ sont semblables à celles de $HC^{\circ}(C,J)$, $HC^{\circ\circ}(C,J)$ et de $HC(C,J)$. Les profils-lignes et les profils-colonnes sont donnés par:

$$F^{\sim c}_{J} = F^{\circ c}_{J} - F^{\circ\circ c}_{J} + F_J$$

et

$$F^{\sim j}_{C} = F^{\circ j}_{C} - F^{\circ\circ j}_{C} + F_C$$

Soit comme origine du nuage de points de R_J, N(C), son centre de gravité F_J, on a :

$$F^{\sim c}_{J} = F^{\circ c}_{J} - F^{\circ\circ c}_{J}$$

et avec comme origine du nuage de points de R_C, N(J), son centre de gravité F_C, on a :

$$F^{\sim j}_{C} = F^{\circ j}_{C} - F^{\circ\circ j}_{C}$$

Les profils s'expriment par:

$$F^{*b}_{J} = F^{b}_{J} - F^{a}_{J} + F_J \text{ et } F^{*j}_{B} = F^{j}_{B} - F_{(BA)}\, F^{j}_{A} + F_B$$

Soit avec comme origine du nuage N(B) de points de R_J son centre de gravité F_J,

on a :

$$F^{*b}_{J} = F^{b}_{J} - F^{a}_{J}$$

et avec comme origine du nuage N(J) de points de R_B son centre de gravité F_B,

on a :

$$F^{*j}_{B} = F^{j}_{B} - F_{(BA)}\, F^{j}_{A}$$

Remarque :

- On a, $\forall c$ et $c' \in b$: $HC^{\sim}_{cj} = HC^{\sim}_{c'j}$
  puisque $\forall c$ et $c' \in b$: $HC^{\circ}_{cj} = HC^{\circ}_{c'j}$ et $HC^{\circ\circ}_{cj} = HC^{\circ\circ}_{c'j}$
  Par le principe d'équivalence distributionnelle, effectuer une analyse des correspondances sur $HC^{\sim}(C,J)$ revient à réaliser celle du tableau $HB^{*}(B,J)$ de terme général :

  $\forall b \in a$: $$HB^{*}_{bj} = HB_{bj} - \frac{p_b}{p_a} HA_{aj} + \frac{p_b}{CC} HC_j$$

  En posant: $$p_b = \frac{HB_b}{CQ} \text{ et } p_a = \frac{HA_a}{CQ}$$

  on retrouve alors le tableau associé à l'inertie intra-classes :

$$HB^{*}_{bj} = HB_{bj} - \frac{HA_{aj}\, HB_b}{HA_a} + \frac{HC_j\, HB_b}{HC}$$

Construction du tableau de "Burt" associé

Sur le même principe que la construction du tableau de Burt associé à l'inertie intra-classes, on a :

$$BB^*_{jj'} = \sum_b \frac{HB^*_{bj}\, HB^*_{bj'}}{HB^*_b}$$

où $$HB^*_{bj} = HB_{bj} - \frac{HA_{aj}\, HB_b}{HA_a} + \frac{HC_j\, HB_b}{HC}$$

on démontre de la même manière que :

$$BB^*_{jj'} = BB_{jj'} - BA_{jj'} + \frac{BB_j BB_{j'}}{CC}$$

Remarque :

- On retrouve l'expression de la décomposition de l'inertie, pour une série de partitions emboîtées, d'après le tableau de Burt où :

  $$BBI_{jj'} = BB^*_{jj'} - \frac{BB_j BB_{j'}}{CC} = BB_{jj'} - BA_{jj'}$$

  En faisant référence à J.J. DAUDIN[1], $BBI_{jj'}$ est la dépendance attachée à B et détachée de A.

L'analyse des correspondances appliquée au tableau associé

Le modèle d'indépendance pour une analyse classique se traduit par un centrage sur l'origine du nuage. Pour une analyse intra-classe, il correspond au centrage des sous-nuages définis par les classes par rapport à leurs centres de gravité.

Ici, on est tout à fait dans une situation qui nous permet d'appliquer une analyse des correspondances sur les tableaux $HC^{\sim}(C,J)$, $HB^*(B,J)$ ou $BB^*(J,J)$ puisque les marges de $HC^{\circ}(C,J)$ et celles du tableau modèle $HC^{\circ\circ}(C,J)$ sont semblables et il en est de même pour les marges des tableaux $HB(B,J)$ et $(\frac{p_B}{p_A})HA(A;J)$ et celles des tableaux de Burt $BB(J,J)$ et $BA(J,J)$. Cette analyse traduit l'analyse intra-classes du nuage $N(B_a)$. De plus par le principe de l'équivalence distributionnelle, les analyses des tableaux $HC^{\sim}(C,J)$, $HB^*(B,J)$ et $BB^*(J,J)$ sont identiques.

Remarques :

- **L'analyse des correspondances du nuage intermédiaire peut être alors considérée comme** une généralisation de l'analyse des correspondances multiples conditionnelle c'est-à-dire **une analyse intra-classes appliquée** non pas à un tableau disjonctif complet mais **à une juxtaposition de tableaux de contingence.** C'est finalement l'analyse intra d'une analyse inter.

- A la différence de l'analyse intra-classes effectuée sur un tableau ternaire, les sous-tableaux définis par les sous-nuages, considérés dans la décomposition de l'inertie pour une série de partitions emboîtées, sont de dimensions différentes.

---

[1][DAUDIN, 1981]

Les formules barycentriques

Pour interpréter l'analyse des correspondances effectuée sur HC~(C,J) ou sur HB*(B,J), nous étudions les formules barycentriques. Notons Φ_Λ les facteurs sur B et Ψ_Λ les facteurs sur J. Ces facteurs sont liés par des formules de transition :

$$\Phi_\Lambda(b) = \frac{1}{\sqrt{\lambda_\Lambda}} \sum_j F^{*b}_j \Psi_\Lambda(j) = \frac{1}{\sqrt{\lambda_\Lambda}} \sum_j (F^b_j - F^a_j) \Psi_\Lambda(j)$$

Une classe b appartenant à a est attirée par les modalités adoptées par les individus c qui la composent. Elle est d'autant plus attirée que l'ensemble des individus composant la classe a possèdent peu cette modalité c'est-à-dire que les autres classes de B_a sont moins représentées par cette modalité.

$$\Psi_\Lambda(j) = \frac{1}{\sqrt{\lambda_\Lambda}}) \sum_b F^{*j}_B \Phi_\Lambda(b) = \frac{1}{\sqrt{\lambda_\Lambda}} \sum_b (F^j_B - \frac{p_b}{p_a} F^j_A) \Phi_\Lambda(b)$$

Comme chaque sous-nuage N(B_a) est recentré à l'origine, la formule de transition se simplifie :

$$\Psi_\Lambda(j) = \frac{1}{\sqrt{\lambda_\Lambda}} \sum_b F^j_B \Phi_\Lambda(b) - \sum_a F^j_A \sum_{b \in a} F_{(ba)} \Phi_\Lambda(b) = \frac{1}{\sqrt{\lambda_\Lambda}} \sum_b F^j_B \Phi_\Lambda(b)$$

On retrouve la formule de transition classique de l'analyse des correspondances. Chaque modalité j est, à $\frac{1}{\sqrt{\lambda_\Lambda}}$ près, au barycentre des classes b de B pondérées par leur importance relative à posséder cette modalité.

Ainsi l'analyse des correspondance du tableau HB*(B,J) dégage les structures à la fois communes aux classes de la partition définie par le niveau intermédiaire, et internes aux classes associées à la partition la moins fine correspondant au niveau supérieur.

Bibliographie

ETUDES URBAINES - COLLECTE DES DONNEES

ASSELIN L.M. (1984) - Techniques de sondage avec applications à l'Afrique, Gaëtan Morin, 697p.

Association pour la Statistique et ses Utilisations (1987) - Les sondages, Economica, 310p.

GHIGLIONE R., MATALON B. (1978) - Les enquêtes sociologiques. Théories et pratique, Armand Colin-collection U, 301p.

COURGEAU D. (1984) - Bases théoriques et modèles pour une enquête sur la biographie familiale. professionnelle et migratoire, Migrations et Urbanisations symposium UGI Rouen.

DESABIE J. (1966) - Théorie et pratiques des sondages, Dunod,477p.

DUBOIS J.L, BLAIZEAU D. (1989) - Connaître les conditions de vie des ménages dans les pays en développement. Tome I : Concevoir l'enquête. Ministère de la coopération et du développement, 165p.

DUREAU F., BARBARY O., MICHEL A., LORTIC B. (1989) - Sondages aréolaires sur image satellite pour des enquêtes socio-démographiques en milieu urbain. Manuel de formation, Orstom, 38p.

GODIN L. (1987) - Préparation des projets urbains d'aménagement, Document technique de la Banque Mondiale, 216p.

GOURIEROUX C. (1981) - Théorie des sondages, Economica, 262p.

Institut National de la statistique et des Etudes Economiques. Institut national d'étude démographique. Ministère de la coopération. Orstom (1977) - Sources et analyse des données démographiques. Applications à l'Afrique d'expression française et à Madagascar. 3ème partie.

LE BRIS E., QUESNEL A., VIMARD P. (1977) - Essai d'enquête spécifique dans une région à forte préssion démographique, Cahiers de l'Orstom, série Sciences Humaines vol 14. n°4, pp.383-408.

LE BRIS E., MARI A., OSMONT A., SINOU A. (1987) - Famille et résidence dans les villes africaines, L'harmattan, 268p.

LE BRIS E., JAGLIN S., MARI A., OSMONT A., OUATTARA A., OUEDRAOGO J.P., PIRON M., SANOU L. (1992) - Les enjeux des extensions urbaines à Ouagadougou, Rapport multigraphié, Orstom-IFU-MRT, 475p.

25ème Congrès de l'Union Géographique Internationale à Paris (août 1984) - De la géographie urbaine à la géographie sociale. Sens et non-sens de l'espace, Collectif français de géographie urbaine et sociale , 262p.

ANALYSE SPATIALE

CHARRE J., MIELLET P., WAINIEZ P., (1991) - Pratique des systèmes d'information géographique raster, Maison de la géographie, coll. Reclus n°18, 50p.

FOTHERINGHAM A.S., WONG D.W.S., (1991) - The modifiable areal unit problem in multivariate statistical analysis, Environment and planing A, vol. 23, pp.1025-1044.

GRASLAND C., (1991) - Espace politiques et dynamiques démographiques en europe, thèse d'université de Paris.

HAGGETT P., (1973))- L'analyse spatiale en géographie humaine, éd. Armand Colin, 390p.

LE GAUFFEY Y., WANIEZ P. (1983) - Stabilité et validité des facteurs sur des données géographiques agrégées : le couple analyse factorielle des correspondances / analyse des tendances, North-Holland, Data Analysis And Informatics, pp119-140.

OPENSHAW S., (1981) - Le problème de l'agrégation spatiale en géographie, L'Espace Géographique n°1, pp.15-24.

PIRON M., (1990) - Structuration de l'information à plusieurs niveaux et analyse des données : application aux données d'une enquête socio-économique à Ouagadougoudou (Burkina Faso), thèse d'université de Paris VI, 265p.

SANDERS L., (1989) - L'analyse des données appliquée à la géographie, coll. Reclus, p224-241.

ANALYSE DES DONNEES

BENZECRI J.P et COLL (1973a) - L'analyse des données. Tome 1: La taxinomie, Dunod, 615p.

BENZECRI J.P et COLL (1973b) - L'analyse des données. Tome 2: L'analyse des correspondances, Dunod, 620p.

BENZECRI J.P (1977a) -Sur l'analyse des tableaux binaires associés à une correspondance. [BIN. MULT.]. Cahiers de l'Analyse des Données. Vol.II. n°1. pp55-71.

BENZECRI J.P (1977b) - Choix des unités et des poids dans un tableau de correspondance. [PONDERATION]. Cahiers de l'Analyse des Données. Vol.II. n°3. pp333-352.

BENZECRI J.P (1979) - Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum. [TAUX QUEST]. Cahiers de l'Analyse des Données. Vol.IV. n°3. pp377-378.

BENZECRI J.P (1980) - Pondérations relatives des questions dans un tableau sous forme disjonctive complète, Cahiers de l'Analyse des Données. vol.V. n°2, p192.

BENZECRI J.P (1982) - Sur la généralisation du tableau de Burt et son analyse par bandes[BANDES BURT], Cahiers de l'Analyse des Données. vol.VII. n°1, pp33-43.

BENZECRI J.P (1983) - Analyse de l'inertie intraclasse par l'analyse d'un tableau de correspondances [ANA.INTRACLASSE], Cahiers de l'Analyse des Données. vol.VIII. n°3, pp351-358.

CAILLIEZ F. et PAGES J.P (1976) - Introduction à l'analyse des données, Smash, 616p.

CAZES P.. BAUMERDER A.. BONNEFOUS S. et PAGES J.P. (1977) - Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives, Cahiers du BURO. n°27, pp3-47.

CAZES P. (1980a) - L'analyse de certains tableaux rectangulaires décomposés en bloc: généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I définitions et applications à l'analyse canonique des variables qualitatives. [ANA. BLOCS 1], Cahiers de l'Analyse des Données. vol.V. n°2, pp145-161.

CAZES P. (1980b) - L'analyse de certains tableaux rectangulaires décomposés en bloc : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. II questionnaires: variantes de codages et de nouveaux calculs de contributions. [ANA. BLOCS 1I], Cahiers de l'Analyse des Données. vol.V. n°4, pp387-403.

CAZES P. CHESSEL D. et DOLEDEC S. (1988) - L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie, Revue de Statistique Appliquée. vol.XXXVI,pp39-54.

DAUDIN J.J. (1981) - Analyse factorielle des dépendances, Revue de Statistique Appliquée. vol.XXIX. n°2, pp15-29.

ESCOFIER B. (1969) - L'analyse factorielle des correspondances, Cahiers du BURO, n°13, pp25-59.

ESCOFIER B. (1978) - Analyse factorielle et distances répondant au principe d'équivalence distributionnelle, Revue de Statistique Appliquée. vol.XXVI. n°4, pp29-37.

ESCOFIER B. (1979) - Stabilité et approximation en analyse factorielle, Thèse de doctorat d'Etat. univ. Paris VI, pp249.

ESCOFIER B. PAGES J. (1982) - Comparaison de groupes de variables définies sur le même ensemble d'individus, Rapports de recherche INRIA n°149, 115p.

ESCOFIER B. et DROUET D. (1983a) - Analyse des différences entre plusieurs tableaux de fréquence [ANA.PLUS.TAB], Cahiers de l'Analyse des Données. volVIII. n°4, pp491-499

ESCOFIER B. (1983b) - Généralisation de l'analyse des correspondances à la comparaison de tableaux de fréquence, Rapports de recherche INRIA n°207, 33p.

ESCOFIER B. (1984) - Analyse factorielle en référence à un modèle. application à l'analyse de tableaux d'échanges, Rapports de recherche INRIA n°337, 19p.

ESCOFIER B. (1987a) - Notice d'utilisation du programme d'analyse factorielle en référence à un modèle, Rapports de recherche INRIA n°616, 13p.

ESCOFIER B. (1987b) - Analyse des correspondances multiples conditionnelle, North-Holland, Data Analysis And Informatics, pp13-22.

ESCOFIER B. et PAGES J. (1988) - Analyses factorielles simples et multiples, Dunod, 241p.

JAMBU M. (1978) - Classification automatique pour l'analyse des données, Dunod, 310p.

LEBART L. (1978) - Programme d'agrégation avec contraintes. [CAH CONTIGUITE], Cahiers de l'Analyse des Données. vol.III. n°3, pp257-287.

LEBART L. MORINEAU A. et TABARD N. (1977) - Techniques de la description statistique, Dunod, 351p.

SAPORTA G. (1975) - Liaisons entre plusieurs ensembles de variables et codage de données qualitatives, Thèse de 3ème cycle. univ Paris VI, 104p.

VOLLE M. (1981) - Analyse des données, Economica. 2ème édition, 317p.

TABLE DES MATIERES

SOMMAIRE 3

INTRODUCTION GENERALE 5

PREMIERE PARTIE : EXEMPLE D'ELABORATION D'UNE STRUCTURE HIERARCHIQUE DE L'INFORMATION 13

Introduction 15

Chapitre 1.1. : STRATEGIE D'ENQUETE 17
- 1.1.1. La dimension temporelle de l'enquête 17
- 1.1.2. Définition de la population statistique 18

Chapitre 1.2. : PREPARATION DE LA COLLECTE ET DE L'ANALYSE DES DONNEES 21
- 1.2.1. Plan de sondage 21
- 1.2.2. Le questionnaire 22
- 1.2.3. Les analyses bivariées 23

Conclusion : VERS L'EXPLOITATION DE LA STRUCTURE HIERARCHIQUE DE L'INFORMATION 29

DEUXIEME PARTIE : CONCEPTION DU TABLEAU POUR EXPLOITER UNE STRUCTURE HIERARCHIQUE DE L'INFORMATION 31

Introduction 33

Chapitre 2.1. : DES DONNEES AUX TABLEAUX 35
- 2.1.1. Quelques définitions et rappels 35
- 2.1.2. Les diverses démarches pour structurer le tableau d'information 38

Chapitre 2.2. : PRINCIPE DE CONSTRUCTION DU TABLEAU D'INFORMATION 41
- 2.2.1. Les conditions 41
- 2.2.2. La relation d'inclusion 43
- 2.2.3. Les relations d'affectation et d'agrégation 45
- 2.2.4. Un système de pondération 46
- 2.2.5. Les niveaux emboîtés 49
- 2.2.6. Niveaux indépendants et généralisation 51
- 2.2.7. Exemple d'une structure complexe 53
- 2.2.8. Synthèse 54

Chapitre 2.3. : LES DIVERS TABLEAUX DU PROCESSUS D'AGREGATION 55
- 2.3.1. Formalisation du processus d'agrégation 55
- 2.3.2. Présentation des tableaux associés aux différents niveaux d'analyse 56
- 2.3.3. Exemple d'interprétation 60

Conclusion : VERS L'ANALYSE D'UN SYSTEME D'ECHELLES 65

TROISIEME PARTIE : UNE METHODE POUR L'ANALYSE D'UN SYSTEME D'ECHELLES 67

Introduction 69

Chapitre 3.1. : LES PRINCIPES DE LA DECOMPOSITION DE L'INERTIE 71
3.1.1. Conditions sur le nuage de points pour une décomposition 71
3.1.2. Décomposition de l'inertie suivant un niveau d'agrégation 72
3.1.3. Décomposition de l'inertie suivant un système d'échelles 74

Chapitre 3.2. : ANALYSE SUR LES ELEMENTS DE LA DECOMPOSITION DE L'INERTIE 79
3.2.1. Rappel sur l'analyse factorielle des correspondances 79
3.2.2. L'analyse inter-classes 81
3.2.3. L'analyse intra-classes 81
3.2.4. L'analyse de "l'inertie intermédiaire" 83
3.2.5. Conséquences de la structure hiérarchique sur les deux nuages 85
3.2.6. Un processus de classification 86

Chapitre 3.3. : AIDES A L'INTERPRETATION DU SYSTEME D'ECHELLES 89
3.3.1. L'analyse des niveaux 89
3.3.2. L'analyse du passage d'un niveau à un autre 91
3.3.3. Critère d'homogénéité des niveaux 92
3.3.4. Niveau d'intervention des variables 96

Conclusion : PROPOSITION D'UNE METHODE D'ANALYSE D'UN SYSTEME D'ECHELLES 99

QUATRIEME PARTIE : UNE APPLICATION DE LA METHODE D'ANALYSE DU SYSTEME D'ECHELLES 101

Introduction 103

Chapitre 4.1. : CONTEXTE EXPERIMENTAL : ENQUETE SUR OUAGADOUGOU 105
4.1.1. Composition du système d'échelles 105
4.1.2. Présentation de la démarche d'analyse 107

Chapitre 4.2. : L'ANALYSE DU SYSTEME D'ECHELLES 111
4.2.1. Analyse au niveau le plus fin : le Ménage 111
4.2.2. Changement d'échelle du niveau Ménage au niveau Parcelle 114
4.2.3. Changement d'échelle du niveau Parcelle au niveau Zone de dénombrement 116
4.2.4. Changement d'échelle du niveau Zone de dénombrement au niveau Secteur 121
4.2.5. Synthèse 126

Chapitre 4.3. : INFLUENCE DU SYSTEME D'ECHELLES 131
4.4.1. Comparaison entre les taux d'inertie 131
4.4.2. Niveau d'intervention des variables 133

Conclusion 137

CONCLUSION GENERALE 141

ANNEXE 1 : TECHNIQUES D'ENQUETE 147

Annexe 1.1 : L'échantillon 149
Annexe 1.2 : Les questionnaires 156
Annexe 1.3 : Déroulement de l'enquête 173
Annexe 1.4 : Quelques principes d'analyse 179
Annexe 1.5 : Le matériel informatique 181

ANNEXE 2 : COMPLEMENTS STATISTIQUES POUR L'ANALYSE DU SYSTEME D'ECHELLES 183

Annexe 2.1 : Les tableaux de "Burt" 185
Annexe 2.2 : La décomposition de l'inertie pour une série de partitions emboîtées 188
Annexe 2.3 : Expression de la décomposition de l'inertie d'après le tableau de Burt 192
Annexe 2.4 : Rappel des principaux travaux sur l'analyse intra-classes 194
Annexe 2.5 : L'analyse inter-classes 195
Annexe 2.6 : L'analyse intra-classes 196
Annexe 2.7 : L'analyse de "l'inertie intermédiaire" 198

BIBLIOGRAPHIE 203

Reproduit par INSTAPRINT S.A.
1-2-3, levée de la Loire – LA RICHE – B.P. 5927 – 37059 TOURS Cedex
Tél. 47 38 16 04

Dépôt légal 2ème trimestre 1992

ORSTOM
Institut français de recherche scientifique pour le développement en coopération
72 route d'Aulnay - 93140 BONDY