



**HAL**  
open science

# L'analyse harmonique qualitative, un outil puissant pour la statistique descriptive des données biographiques

Olivier Barbary

► **To cite this version:**

Olivier Barbary. L'analyse harmonique qualitative, un outil puissant pour la statistique descriptive des données biographiques. XXVIIIe journées de statistiques de L'ASSU, 27-30 mai 1996, Université Laval - Québec, May 1996, Quebec, Canada. hal-02481360

**HAL Id: hal-02481360**

**<https://hal.science/hal-02481360v1>**

Submitted on 17 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Olivier Barbary  
Statisticien IRD,

Departamento de matemáticas y estadística  
Universidad Nacional de Colombia - Santafé de Bogotá D.C., Colombia  
tel/fax : (57 2) 884 52 11, email : [negritud@mafalda.univalle.edu.co](mailto:negritud@mafalda.univalle.edu.co)

XXVIII<sup>e</sup> journées de statistiques, 27-30 mai 1996, Université Laval - Québec

## **L'ANALYSE HARMONIQUE QUALITATIVE, UN OUTIL PUISSANT POUR LA STATISTIQUE DESCRIPTIVE DES DONNÉES BIOGRAPHIQUES**

### **INTRODUCTION**

La collecte de données biographiques est de plus en plus fréquente en sciences sociales. Dans l'étude des formes de la mobilité humaine, elle tente de répondre à des questionnements nouveaux, dans des contextes thématiques, disciplinaires et méthodologiques variés. Qu'il s'agisse en effet de la mobilité spatiale, sociale ou économique, de celle des individus ou des groupes sociaux, d'approches démographiques, géographiques, économiques ou socio-anthropologiques, de méthodes quantitatives ou qualitatives, l'objet de la recherche est à chaque fois un phénomène continu (les vies humaines sont composées d'une infinité d'instant, de lieux, d'événements...). Dès lors, même si aucun système d'observation ne peut prétendre rendre compte de cette continuité "théorique", l'appareil des concepts et des méthodes de collecte doit viser l'observation la plus exhaustive possible du temps et des espaces dans lesquels se déroule la mobilité des hommes. Cette réflexion n'en est plus aujourd'hui au stade des débats théoriques, du moins à ce niveau de généralité. Récemment, les systèmes d'observation statistique de la mobilité - entendons par là le domaine qui s'étend de la définition des populations et des concepts de base de l'observation, à la réalisation d'enquêtes sur des échantillons représentatifs - ont connu de grands progrès.

Mais si l'on sait de plus en plus précisément pourquoi et comment collecter des données biographiques quantitatives, de nombreuses difficultés subsistent dans leur analyse. On doit en effet employer des méthodes statistiques respectant la richesse et la continuité du corpus. De l'avis général, il existe à l'heure actuelle un décalage important entre l'évolution des concepts et des techniques de collectes, qui ont permis une avancée significative dans la saisie quantitative des réalités complexes de la mobilité des populations, et le caractère encore très frustré des outils de statistique descriptive généralement employés lors de l'exploitation. On s'en tient souvent à des analyses purement transversales et, lorsque ce n'est pas le cas, à l'examen d'indicateurs longitudinaux monovariés, nécessairement réducteurs de l'information originale. Globalement, cet état de fait a pour conséquence une sous-exploitation des enquêtes; la pratique de l'analyse multivariée des données biographiques doit elle aussi se généraliser.

Il faut dire à la décharge des équipes de chercheurs confrontés à ce problème, qu'à cause de la structure complexe des données, les résultats applicables de la recherche statistique sur ces questions sont assez récents. Généralement formalisée à l'aide de la théorie des processus stochastiques, l'analyse statistique des données biographiques sur gros échantillons est un domaine resté longtemps assez théorique ou d'application très spécifique. En particulier, à la suite des travaux de Cox [1972] sur l'analyse des données de survie, les approches statistiques des données longitudinales sont le plus souvent cantonnées dans le domaine de l'inférence et de la modélisation (voir par exemple les travaux de Courgeau et Lelievre [1989] à l'INED ou l'ouvrage de synthèse publié sur ce thème par Droesbeke, Fichet et Tassi [1989]). Aujourd'hui cependant, les acquis méthodologiques sur la statistique descriptive multivariée des données biographiques se sont diversifiés et deviennent très prometteurs. C'est ce que nous tenterons de montrer dans cet exposé, en partant d'une technique déjà éprouvée, l'analyse harmonique qualitative (Deville et Saporta [1980]), que nous avons complétée pour l'appliquer à l'étude de la mobilité spatiale des populations de Bogota.

L'enquête et le travail d'exploitation statistique présentés ici s'inscrivent dans une recherche entreprise depuis août 1992 par une équipe franco-colombienne de chercheurs du CEDE (Centro de Estudio sobre el Desarrollo Económico, C.E. Florez) et de l'ORSTOM (F. Dureau et M.C. Hoyos). Ce programme porte sur la mobilité des populations de Bogota et son impact sur la dynamique de l'aire métropolitaine (enquête CEDE/ORSTOM-1993). En parallèle à ces recherches de sciences sociales, se développe depuis février 1994 un second programme de coopération entre l'Université nationale de Colombie (C.E. Pardo et J. Ramos ) et l'ORSTOM (O. Barbary) qui porte sur les méthodes statistiques d'analyse des données longitudinales.

Le questionnaire de l'enquête comprend un formulaire d'information socio-démographique couvrant l'ensemble des individus des ménages sélectionnés dans l'échantillon, ainsi qu'une série de modules visant à recueillir, sous forme de calendriers, des données rétrospectives sur la biographie résidentielle, professionnelle et familiale d'un sous-échantillon de la population des ménages appelé "échantillon biographique". L'objectif de la première phase d'analyse de ces données longitudinales est de réaliser une "mise à plat", la plus complète possible, des éléments biographiques recueillis sur les 1031 individus de l'échantillon en orientant la description sur le thème de la mobilité résidentielle au sein de l'agglomération.

Dans la première partie, nous décrivons d'abord rapidement l'ensemble des données utilisées pour cette analyse : définition de la variable active et des variables illustratives. Puis nous donnons les principes de la méthode d'analyse typologique sur laquelle se base l'exploitation : l'analyse harmonique qualitative associée à une technique de classification automatique. Nous exposons ensuite comment la méthode a été mise en oeuvre sur les données pour parvenir à une typologie en quinze classes qui constitue le résultat principal.

Dans le temps imparti nous ne pouvons malheureusement pas aborder les résultats détaillés pour les 15 classes. La deuxième partie vise simplement une illustration de la démarche. On décrit succinctement le schéma général de la mobilité spatiale à l'intérieur de l'agglomération de Bogota que permet de construire la typologie. Puis on s'intéresse aux relations existantes entre cette mobilité et les autres composantes de la biographie à partir de deux exemples où apparaissent les modalités particulières de l'insertion résidentielle et professionnelle des jeunes migrants.

## **1. DONNÉES ET MÉTHODE**

### **1.1. Les données**

Bogota est la métropole latino-américaine qui a connu la croissance démographique la plus rapide durant les années cinquante et soixante (plus de 6% par an). Entre 1951 et 1964 la ville a doublé sa population et en 1970 elle comptait 2,5 millions d'habitants. Depuis une vingtaine d'années, le rythme de croissance de la capitale colombienne, comme celui des autres métropoles du sous-continent, s'est ralenti : il était d'environ 2,5% annuel lors du dernier recensement en 1985. Au moment de l'enquête CEDE/ORTOM (fin 1993), Bogota compte près de 5,5 millions d'habitants et croit toujours à un rythme légèrement supérieur à 2% par an. Ce ralentissement de la croissance s'explique par les effets conjugués de trois phénomènes : la réduction du rythme d'accroissement naturel due à la baisse rapide de la fécondité, une diminution des flux migratoires à destination de Bogota proprement dit et la transformation du schéma géographique de la croissance au profit des municipalités périphériques de l'aire métropolitaine. L'évolution démographique récente de la ville s'accompagne de nouvelles stratégies de localisation résidentielle des habitants qui produisent des changements rapides et importants dans la répartition de la population et les formes de ségrégation au sein de l'agglomération. Ce sont ces recompositions, mal mesurées et peu étudiées jusqu'à présent, qui sont au centre de la thématique de l'enquête.

Le plan de sondage adopté ne vise pas une représentativité de l'ensemble de l'aire métropolitaine de Bogota mais est adapté à l'observation fiable de 11 zones d'études (4 communes ou parties de communes de l'aire métropolitaine et 7 quartiers de Bogota). Ces zones ont été choisies *a priori* comme ayant une valeur euristique pour la problématique du programme. Le sous-échantillon auquel est soumis le questionnaire biographique est quant à lui composé d'individus de plus de 18 ans (un

individu par ménage enquêté) et sa structure contrôlée par des quotas de sexe, d'âge, de relation de parenté avec le chef de ménage et de statut migratoire. Les conclusions de l'analyse n'ont donc de portée qu'à l'intérieur de cet univers, pour sa population particulière, et non pour l'ensemble de la population de Bogota et de son aire métropolitaine.

Visant en premier lieu une typologie des trajectoires spatiales parcourues par les individus en ville, nous avons choisi de baser l'analyse sur la variable d'état construite à partir de l'observation des localisations dans l'aire métropolitaine de l'ensemble des résidences de plus d'un an connues par les individus (variable active). Le niveau choisi pour l'agrégation géographique de la variable doit fournir la précision maximum dans l'analyse de la mobilité spatiale intra-urbaine tout en conservant des effectifs suffisants dans chaque modalité. La nomenclature géographique qui convient le mieux à cet objectif est celle des "alcaldas menores" de Bogota : 19 unités que nous appellerons désormais arrondissements. À ces modalités décrivant les résidences à l'intérieur de la ville s'ajoute une modalité pour les lieux de Bogota non spécifiés, quatre modalités pour les quatre zones d'étude de la périphérie de l'aire métropolitaine (communes de Chia, Tabio, Madrid, Soacha), une modalité pour les autres communes de l'aire métropolitaine ( autres mun. a.m.) et enfin une modalité pour les résidences situées hors de l'a.m.

Pour mettre à jour les relations existant entre les types de mobilité spatiale et les autres composantes de la biographie, nous introduisons un certain nombre de variables illustratives longitudinales. Les 7 variables retenues résument les "chapitres" de la biographie que l'on peut supposer être déterminés et/ou déterminants par rapport à la trajectoire spatiale : étapes du cycle de vie (relation de parenté avec le chef de ménage, statut matrimonial, co-résidence avec les enfants et composition du ménage), accès au logement, carrière d'éducation et mobilité professionnelle. Chaque variable ayant son propre calendrier, les nombres d'étapes vécues par les individus de l'échantillon sont donc différents pour chacune d'elles, et indépendants du nombre d'étapes résidentielles.

Enfin une caractérisation socio-économique des individus pratiquant chaque type de mobilité et des ménages auxquels ils appartiennent sera obtenue grâce à un second ensemble de variables illustratives transversales, caractéristiques individuelles et collectives à la date de l'enquête. On reprend, dans ce bloc de variables, les descripteurs socio-démographiques classiques des individus -sexe, âge, statut migratoire, niveau d'instruction, catégorie socioprofessionnelle etc, et des ménages -taille du ménage, caractéristiques du logement (taille, statut d'occupation, indice de promiscuité) et du chef de ménage (sexe, âge, statut migratoire, âge moyen des enfants).

*La variable longitudinale active : RESIALC*

Tableau 1 : Fréquences selon l'arrondissement de résidence (étapes vécues par les individus de l'échantillon biographique)

RESIALC			Cumul	
	Freq.	Freq. (%)	Freq.	Freq. (%)
.1 Usaquen	289	5.9	289	5.9
2 Chapinero	234	4.7	523	10.6
3 Santa fe	455	9.2	978	19.8
4 San Cristob	62	1.3	1040	21.1
5 Usme	25	0.5	1065	21.6
6 Tunjuelito	65	1.3	1130	22.9
7 Bosa	174	3.5	1304	26.4
8 Kennedy	127	2.6	1431	29.0
9 Fontibón	89	1.8	1520	30.8
10 Engativa	232	4.7	1752	35.5
11 Suba	64	1.3	1816	36.8
12 Barrios Uni	144	2.9	1960	39.7
13 Teusaquillo	115	2.3	2075	42.0
14 Los Martires	80	1.6	2155	43.6
15 Antonio Nariño	147	3.0	2302	46.6
16 Puente Arenda	111	2.2	2413	48.9
17 La Candelaria	117	2.4	2530	51.2
18 Rafael Uribe	258	5.2	2788	56.5
19 Ciudad Bolivar	46	0.9	2834	57.4
20 Bog. non spec.	32	0.6	2866	58.1
21 Chia	205	4.2	3071	62.2
22 Tabio	122	2.5	3193	64.7
23 Madrid	202	4.1	3395	68.8
24 Soacha	144	2.9	3539	71.7
25 Autres Mun AM	163	3.3	3702	75.0
26 Hors AM	1235	25.0	4937	100.0

*Les variables longitudinales illustratives*

Tableau 2 : Fréquences selon la relation de parenté avec le chef de ménage (étapes vécues par les individus de l'échantillon biographique)

<b>Cumul</b>	<b>Cumul</b>		<b>Freq.</b>	<b>Freq. (%)</b>
<b>PARENJH</b>				
<b>Freq.</b>	<b>Freq. (%)</b>			
<b>1 Chef du ménage</b>			581	20.8
581	20.8			
<b>2 Conjoint du c.m.</b>			478	17.1
1059	37.9			
<b>3 Enfant du c.m.</b>			1191	42.6
2250	80.5			
<b>4 Autre parent du c.m.</b>			295	10.6
2545	91.0			
<b>5 Autre non parent du c.m.</b>			184	6.6
2729	97.6			
<b>6 Ménages collectif</b>			67	2.4
2796	100.0			

A travers la relation de parenté qui lie l'individu enquêté aux chefs des ménages dans lesquels il a successivement vécu, on rend compte de l'évolution du statut de l'individu dans le ménage selon les 6 modalités du tableau 2.

Tableau 3 : Fréquences selon le statut matrimonial (étapes vécues par les individus de l'échantillon biographique)

<b>Cumul</b>				<b>Cumul</b>
<b>ESTCIV</b>		<b>Freq.</b>	<b>Freq. (%)</b>	<b>Freq.</b>
<b>Freq. (%)</b>				
<b>1 Célibataires</b>		1031	49.6	1031
49.6				
<b>2 Vie en union</b>		837	40.2	1868
89.8				
<b>3 Séparés</b>		155	7.5	2023
97.3				
<b>4 Veufs</b>		57	2.7	2080
100.0				

Dans la succession des situations matrimoniales, on prend en compte toutes les unions sans distinguer leur caractère légal ou non (tableau 3).

Tableau 4 : Fréquences selon le nombre d'enfants co-résidant avec l'individu (étapes vécues par les individus de l'échantillon biographique)

Cumul NUMHIJ Freq. (%)		Freq.	Freq. (%)	Cumul Freq.
-----				
23.2	<b>1 1 enfant</b>	822	23.2	822
42.3	<b>2 2 enfants</b>	675	19.1	1497
54.9	<b>3 3 enfants</b>	447	12.6	1944
62.1	<b>4 4 enfants</b>	254	7.2	2198
66.4	<b>5 5 enfants</b>	154	4.3	2352
68.3	<b>6 6 enfants et+</b>	66	1.9	2418
100.0	<b>7 sans enfants</b>	1123	31.7	3541

Plutôt que l'évolution de la descendance de l'individu, nous avons choisi de retenir le nombre de ses enfants avec lesquels il a cohabité au cours de sa vie, plus directement en relation avec la stratégie résidentielle (tableau 4).

Tableau 5 : Fréquences selon la composition du ménage (étapes vécues par les individus de l'échantillon biographique)

Cumul Freq.	Cumul COMPLOG Freq. (%)		Freq.	Freq. (%)
1551	52.9	<b>1 Nucléaire complet</b>	1551	52.9
2119	72.2	<b>2 Nucléaire incomplet</b>	568	19.4
2333	79.5	<b>3 Étendu complet</b>	214	7.3
2368	80.7	<b>4 Étendu incomplet</b>	35	1.2
2934	100.0	<b>5 Autres types</b>	566	19.3

Pour compléter la description de la constitution de la famille des l'enquêtés, on a retenu (tableau 5) une version simplifiée de la typologie de la composition des ménages utilisée dans l'étude de C.E. Florez et R. Mendez [1994].

Tableau 6 : Fréquences selon le statut d'occupation du logement (étapes vécues par les individus de l'échantillon biographique)

Cumul Freq.	Cumul OCUPVIV Freq. (%)		Freq.	Freq. (%)
1367	48.2	<b>1 Prop. d'un membre du ménage</b>	1367	48.2
2399	84.6	<b>2 Location</b>	1032	36.4
2746	96.8	<b>3 Usufruit</b>	347	12.2
2836	100.0	<b>4 Autres</b>	90	3.2

L'accès au logement de l'individu est appréhendé à travers le statut sous lequel le ménage dans lequel il vit occupe le logement, en quatre modalités (tableau 6)



Tableau 7 : Fréquences selon le niveau d'étude en cours  
(étapes vécues par les individus de l'échantillon biographique)

<b>Cumul</b>				<b>Cumul</b>
<b>NIVEDUC</b>		<b>Freq.</b>	<b>Freq. (%)</b>	<b>Freq.</b>
<b>Freq. (%)</b>				
-----				
	<b>1 N'étudie pas</b>	2116	48.9	2116
48.9				
	<b>2 Maternel ou primaire</b>	1179	27.3	3295
76.2				
	<b>3 Secondaire</b>	674	15.6	3969
91.8				
	<b>4 Supérieur</b>	240	5.6	4209
97.4				
	<b>5 Autre type d'enseignement</b>	114	2.6	4323
100.0				

La variable décrit classiquement les différentes étapes en cours de la scolarisation des individus; durant les périodes où l'individu n'étudie pas, la variable prend la modalité 1 (tableau 7).

Tableau 8 : Fréquences selon la catégorie socio-professionnelle  
(étapes vécues par les individus de l'échantillon biographique)

Cumul				C
CSO	Freq.	Freq. (%)	Freq.	Freq. (%)
Freq.	Freq. (%)			
-----				
<b>1 Ne travaille pas</b>			1372	37.6
1372	37.6			
<b>2 Directeurs</b>			22	0.6
1394	38.2			
<b>3 Fonctions spécifiques*</b>			80	2.2
1474	40.4			
<b>4 Prop. de petites entrep.</b>			30	0.8
1504	41.3			
<b>5 Petits producteurs indépendants</b>			412	11.3
1916	52.6			
<b>6 Cadres et techniciens</b>			203	5.6
2119	58.1			
<b>7 Employés</b>			304	8.3
2423	66.5			
<b>8 Ouvriers qualifiés</b>			853	23.4
3276	89.9			
<b>9 Ouvriers non qualifiés</b>			96	2.6
3372	92.5			
<b>10 Employés domestiques</b>			270	7.4
3642	99.9			
<b>11 Non spécifiée</b>			4	0.1
3646	100.0			
* : En Colombie on désigne sous cette appellation les professions dite "libérales" en France et les cadres supérieurs, sans distinction de statut professionnel : ingénieurs, médecins, architectes, juristes, économistes etc, qu'ils soient patrons, salariés ou à leur compte)				

En l'absence d'une classification des catégories socioprofessionnelles faisant référence dans le système statistique colombien, la typologie des emplois retenue pour résumer les carrières professionnelles des individus s'inspire de celle développée en Argentine par M. Arino, M.E. Rosas et S. Torrado [1987].

## 1.2. La méthode

L'analyse harmonique qualitative (AHQ) a été présentée pour la première fois dans un court article théorique par Deville et Saporta en 1980, puis à nouveau plus longuement par Deville, dans un article de 1982, où elle est appliquée à des données sur la nupcialité d'une cohorte de femme ayant connue plusieurs unions. Avec cette proposition, l'analyse des correspondances (Benzecri [1973]) trouve un nouveau terrain d'application : la statistique descriptive et typologique des données longitudinales. Sans aborder la justification mathématique assez complexe de la méthode, ce qui serait trop long ici, il convient de dire que l'on peut adopter plusieurs solutions d'approximation du problème complet par recodage de l'information de départ. La plus simple et efficace (retenue pour ce travail) est celle de la division de l'intervalle de temps de l'observation en un nombre raisonnable de périodes (de durée constante ou non) sans tenir compte des changements d'états individuels. Un tableau de fréquence est ensuite construit en calculant la proportion de temps passé par les individus dans chacun des états possibles des variables longitudinales au cours de chacune des périodes de recodage. Ce tableau est soumis à l'analyse des correspondances.

Telle quelle la technique reçoit sans aucun doute, selon l'expression de Deville [1982], son "brevet de navigabilité"; en effet sa justification théorique est solide et son champs potentiel d'application extrêmement vaste. Cependant, comme en témoigne Beret [1988] dans un article où il analyse des données sur l'insertion professionnelle des jeunes, l'interprétation directe des facteurs devient difficile et parfois décevante lorsque l'information biographique est détaillée. Cette complexité de l'interprétation, jointe au fait qu'il n'existe pas jusqu'à présent de logiciel proposant une mise en oeuvre simple et paramétrée de la méthode, explique probablement les limites actuelles de son application aux données d'enquêtes. Pourtant, c'est sans doute la méthode d'analyse descriptive des processus la plus puissante et la plus souple existant à l'heure actuelle.

Pour construire un outil complet et d'un emploi aisé d'analyse de données longitudinales basé sur l'AHQ, nous avons élaboré sous le logiciel SAS un programme paramétré qui complète la méthode factorielle originale par un processus de classification. Le programme effectue selon le choix de l'utilisateur différents types de recodages de données longitudinales pour l'analyse des correspondances; puis il enchaîne l'analyse factorielle du tableau recodé, la classification des itinéraires individuels dans l'espace des facteurs et la caractérisation des classes. Pour mener une interprétation approfondie, on peut repartir du tableau de fréquences et faire appel à un logiciel spécialisé (SPADN) pour l'analyse des correspondances, la classification et leur interprétation.

## 1.3. La mise en oeuvre sur les données

### *Le temps de l'analyse*

L'analyse des données présentées dans la section précédente va être menée selon ce qu'on peut appeler le temps biographique individuel, c'est à dire en suivant les individus depuis leur naissance jusqu'à l'âge atteint à la date de l'enquête. D'autres options sont possibles, on pourrait par exemple mener l'analyse en temps historique (en suivant les individus entre deux dates) ou encore selon une horloge démarant à un événement autre que la naissance. Mais le choix du temps biographique correspond à l'orientation problématique de l'exploitation des données. En effet il permet seul de mettre en relation statistique, dans l'ensemble de l'échantillon ou dans chacune des classes de la typologie, les différents itinéraires résidentiels, familiaux, éducatifs ou professionnels, pour que s'exprime la cohérence des stratégies individuelles.

### *La période d'analyse*

L'option du temps biographique étant prise, il faut recoder les données et donc choisir une période de temps commune à tous les individus, quelque soit leur âge à la date de l'enquête. Si l'on souhaite conserver la totalité des étapes vécues, c'est l'âge atteint à la date de l'enquête par l'individu le plus vieux (92 ans) qui fixe l'étendue de la période. Afin d'éviter que la partie finale du tableau soit presque totalement vide, on préfère arrêter l'analyse à 65 ans pour la variable active (70 ans pour les variables illustratives sauf la carrière éducative arrêtée dès 45 ans).

### *Les données censurées*

Les individus n'ayant pas atteint cet âge sortent d'observation, donc des modalités prévues pour les variables longitudinales, à partir de leur âge à la date de l'enquête (phénomène appelé "censure à droite" dans le jargon de l'analyse longitudinale). On note au passage que le type de censure que l'on doit prendre en compte dépend du type de temps choisis pour l'analyse : avec un temps calendaire par exemple, les données seraient "censurées à gauche" pour les individus n'étant pas encore nés à une date donnée. La solution retenue pour cette première analyse sur l'ensemble de l'échantillon est l'ajout d'une modalité supplémentaire à chacune des variables longitudinales, modalité dans laquelle l'individu entre dès qu'il est censuré. Ainsi l'ensemble des individus de l'échantillon sont "présents" dans le tableau tout au long de la période d'analyse (de 0 à 65 ans).

### *Découpage de la période d'analyse et calcul des fréquences*

Comme on l'a dit en présentant la méthode, la mise en oeuvre de l'AHQ repose sur un découpage de la période d'analyse en un nombre "raisonnable" de périodes de recodage. Une fois ce découpage réalisé, deux solutions s'offrent pour construire le tableau soumis à l'AFC :

- calculer pour chaque individu et chaque période de recodage, la proportion du temps total de l'analyse passé dans chaque état : la métrique est uniforme sur le temps, la somme de chaque ligne vaut 1;
- calculer la proportion du temps de chaque période passée dans chaque état : la métrique sur le temps est non-uniforme si les périodes de recodage n'ont pas une durée constante, la somme de chaque ligne est égale au nombre de périodes définies.

Du point de vue du calcul algébrique effectué lors de l'analyse factorielle, rien n'oblige à ce que les différentes périodes de recodage soient de même durée ou que la métrique sur le temps soit uniforme. Bien au contraire d'une part plusieurs arguments, intuitifs et statistiques, militent en faveur d'un découpage plus détaillé dans les périodes où les changements d'état sont nombreux (voir J.C. DEVILLE [1982] et A. Florette [1988]). D'autre part, du point de vue de la problématique qui nous occupe (stratégie résidentielle des individus), il est naturel de s'intéresser plus spécialement aux changements de résidences qui surviennent à l'âge adulte, où la responsabilité de la décision revient dans la plupart des cas à l'individu, plutôt qu'aux changements de résidences antérieurs ou postérieurs, plus souvent décidés par des tiers.

Nous avons tout d'abord testé un découpage uniforme de la période 0-65ans en 13 segments quinquennaux, les fréquences étant calculées en proportion du temps total de la période analysée. Les résultats de cette première analyse (typologie en 22 classes) sont d'une part excessivement dominés par l'effet de l'âge à la date de l'enquête. D'autre part les classes regroupant les individus jeunes sont peu homogènes et à l'inverse la description des groupes de population âgée est trop détaillée. Nous devons alors refaire deux analyses spécifiques sur l'ensemble des classes jeunes et celui des classes âgées.

L'examen de l'histogramme de distribution des changements d'état selon l'âge des individus (après exclusion des 1031 situations à la naissance on passe à 2856 changements d'état, voir figure 1 en annexe) permet de définir un découpage mieux adapté aux données en 15 périodes d'amplitude variable correspondant approximativement aux quantiles de la distribution : la fréquence d'occurrence des changements d'états dans chaque période varie de 4% à 8% du total. La précision est forte entre 13 et 25 ans, elle diminue avant et après. Une démarche semblable a été appliquée pour les variables longitudinales de statut matrimonial, de co-résidence avec les enfants et de niveau d'éducation. Les autres variables illustratives longitudinales ont été recodées selon 12 périodes quinquennales de 0 à 59 ans et une treizième période de 60 à 70 ans. Pour l'ensemble des variables, les fréquences du tableau soumis à l'analyse des correspondances sont calculées en proportion du temps de chaque période de recodage.

### *L'analyse factorielle*

Après suppression des colonnes entièrement nulles (pas d'individus dans l'état au cours de la période), le tableau final recodé comprend 1031 lignes, 398 colonnes actives et 625 colonnes illustratives. L'analyse des correspondances (PROC CORRESP de SAS ou PROC CORBI de SPADN) fournit un histogramme de valeurs propres très plat (figure 2). Ceci ne doit pas surprendre étant donné la structure du tableau : grand nombre de colonnes au regard du nombre de lignes et abondance de colonnes presque vides. Il faut considérer ici que le recodage adopté, s'il conserve dans leur précision temporelle originale la totalité des durées de séjours des individus dans les états, sacrifie en revanche

l'ordre chronologique des étapes puisque toute permutation des colonnes est indifférente pour le résultat de l'AFC. De ce fait, deux individus ayant des itinéraires strictement semblables mais simplement décalés d'une ou deux années, apparaissent très différents. Le tableau final tel qu'il est analysé peut donc être considéré comme très "bruité" par rapport à la structure des proximités qui nous intéressent, "bruitage" que l'on retrouve dans l'allure de l'histogramme des valeurs propres.

**Figure 2 :** *Histogramme des premières valeurs propres de l'AFC du tableau 1031x398*

-----  
 SPAD.N Sistema Portable para el Analisis de Datos  
 Copyright (C) CISIA, 1987, 1991 - Version 2.52  
 -----

```

+-----+-----+-----+-----+
+-----+
-----
|NUMERO |VALOR      |PORCENTA. |PORCENTA |
|        |PROPIO     |           |ACUMU.   |
+-----+-----+-----+-----+
+-----+
-----
|  1  | .7099  |  3.17  |  3.17  |
*****
|  2  | .6291  |  2.81  |  5.98  |
*****
|  3  | .5973  |  2.67  |  8.65  |
*****
|  4  | .5800  |  2.59  | 11.24  |
*****
|  5  | .5562  |  2.48  | 13.72  |
*****
|  6  | .5291  |  2.36  | 16.09  |
*****
|  7  | .5046  |  2.25  | 18.34  |
*****
|  8  | .4848  |  2.17  | 20.51  |
*****
|  9  | .4834  |  2.16  | 22.67  |
*****
| 10  | .4602  |  2.06  | 24.72 |
*****
| 11  | .4562  |  2.04  | 26.76  |
*****
| 12  | .4392  |  1.96  | 28.72  |
*****
| 13  | .4266  |  1.91  | 30.63  |
*****
| 14  | .4239  |  1.89  | 32.52  |
*****
| 15  | .4197  |  1.87  | 34.40  |
*****
| 16  | .4073  |  1.82  | 36.22  |
*****
| 17  | .3915  |  1.75  | 37.97  |
*****
| 18  | .3903  |  1.74  | 39.71  |
*****
| 19  | .3869  |  1.73  | 41.44  |
*****

```

etc...

Malgré cela, l'interprétation des premiers facteurs est assez aisée et, s'il est vrai que la chronologie est perdue lors du recodage, elle structure si fortement les données (méta-information) qu'elle ressort sur tous les axes factoriels utiles. Par exemple les trois premiers facteurs (9% environ de l'inertie du nuage), sont presque entièrement déterminés par les groupes de variables afférentes aux modalités 22 (commune de Tabio, la séquence T1A22 à T15A22 totalise 94% de l'inertie du premier facteur), 21 (commune de Chia, 56% de l'inertie du deuxième facteur et 30,2% de l'inertie du troisième) et 23 (commune de Madrid, 11,2% de l'inertie du second facteur et 60% de l'inertie du troisième). Le plan 1x2 (figure 3) montre le type de structure mise en évidence par ces premiers axes : les séquences de variables correspondant à une même modalité sont regroupées, ordonnées chronologiquement le long des axes, elles correspondent à des sous ensembles d'individus minoritaires mais très fortement caractérisés par chaque groupe de variables. Il s'agit, on l'aura deviné, des sous-groupes de l'échantillon stables dans ces modalités : dans notre cas, trois groupes ayant vécu toute leur vie respectivement à Tabio, Chia et Madrid. Pour compléter la démarche classique de l'analyse des correspondances, on peut ensuite projeter sur le plan les variables illustratives longitudinales ou faire apparaître le point représentant le centre de gravité d'une caractéristique transversale. En appliquant cette démarche à l'ensemble des facteurs "interprétables", on parvient à identifier et caractériser l'ensemble des groupes stables de l'échantillon sans difficulté. Ces individus stables représentent à peu près 22% de l'échantillon (voir plus bas le résultat de la classification) et environ 52% de l'inertie expliquée par les 10 premiers facteurs. L'on peut, sur des critères empiriques, retenir ces 10 axes comme les facteurs "utiles", ils totalisent 24.7% de l'inertie totale, le reste étant considéré comme du bruit.

Les schémas de mobilité qui caractérisent le reste des individus (donc la grande majorité de l'échantillon!) sont moins faciles à mettre en évidence à partir de la seule interprétation des facteurs. Ce qui caractérise les axes, ce sont des associations entre modalités qui témoignent de transitions fréquentes, à certains âges, entre les lieux qu'elles représentent. Nous n'avons plus, comme dans le cas de la stabilité, d'axes entièrement déterminés par un ou deux groupes d'individus aux trajectoires globalement semblables, mais seulement la mise en évidence d'un groupe ayant en commun une transition donnée à un âge donné. Ce genre de résultat est d'ailleurs typique de l'application de l'analyse harmonique qualitative aux données de calendriers (voir Deville [1982], P. Béret [1988] [1995]), mais du point de vue de notre objectif typologique, il est insuffisant. Pour parvenir à une première typologie, nous avons donc procédé à une classification des individus dans l'espace vectoriel des premiers facteurs de l'analyse des correspondances <sup>1</sup>. Finalement l'analyse des correspondances a pour principal intérêt, mais c'est déjà beaucoup, d'éliminer de l'information de départ la majorité du bruit.

---

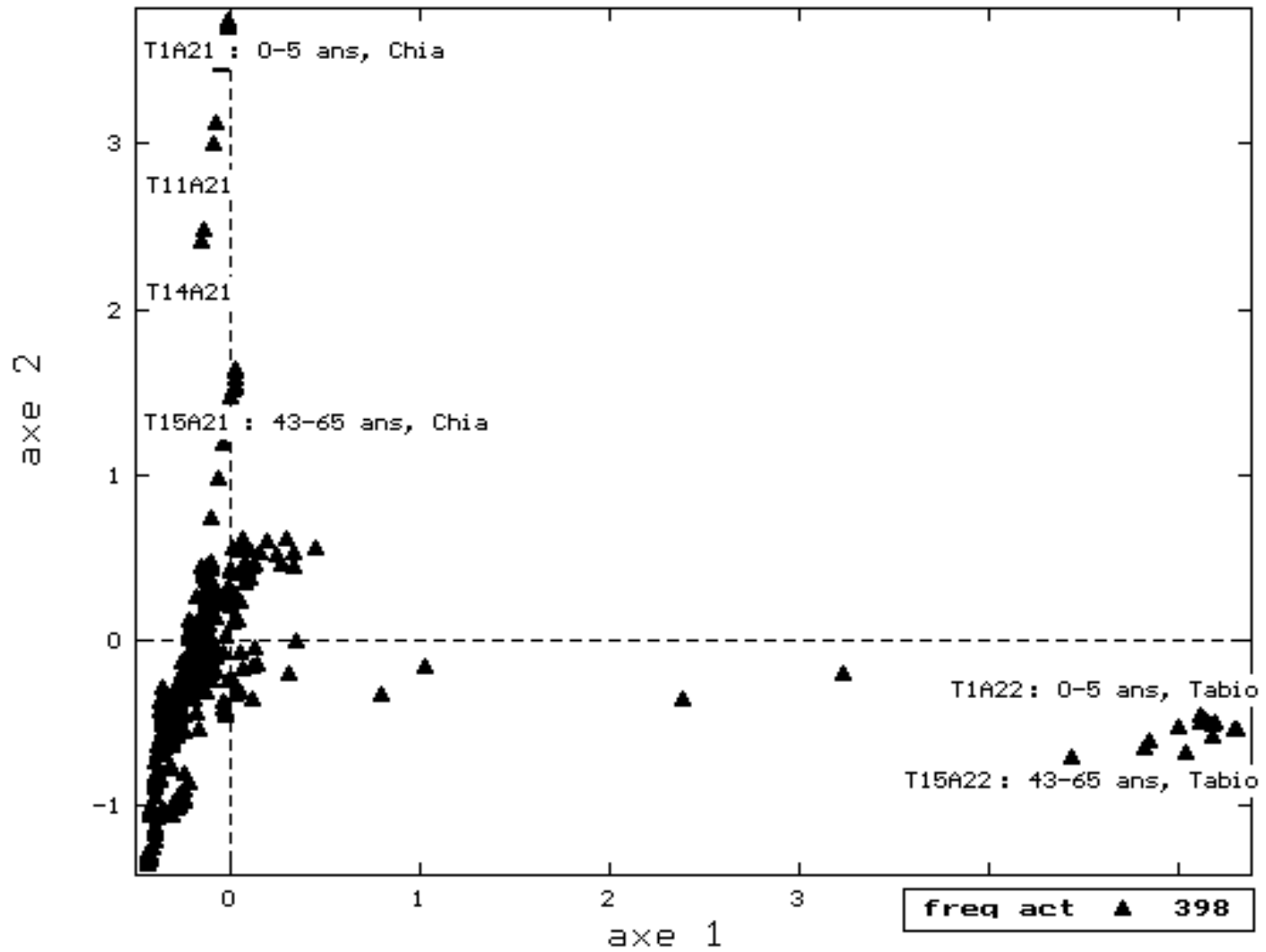
<sup>1</sup> : C'est cette même démarche qu'appliquent A. Degenne, M. O. Lebeaux, L. Mounier [1995] aux données sur l'insertion professionnelle des jeunes.





Figure 3 : Nuage des variables longitudinales actives sur le plan factoriel 1x2

Fichier: P12ACT.FGS





### *La classification*

Après plusieurs essais, en faisant varier le nombre de facteurs de 7 à 15 et en explorant les partitions jusqu'à 30 classes et plus, on constate que jusqu'à 10 facteurs la typologie gagne en précision puis, à partir du onzième facteur, la taille de la classe la plus importante augmente, même si l'on considère des partitions comprenant un grand nombre de classes dont beaucoup ont par conséquent des effectifs trop faibles. Sur cette base empirique, nous avons donc retenu l'espace vectoriel constitué par les dix premiers facteurs de l'AFC, et soumis le nuage des individus dans cet espace à des algorithmes de classification ascendante hiérarchique ou semi-hiérarchique (PROC CLUSTER - critère de Ward sous SAS, PROC SEMIS de SPADN). Les deux procédures donnent des résultats très proches et nous avons conservés ceux de SPADN qui présentent l'avantage d'optimiser la partition une fois choisi le nombre de classes (figure 4).

**Figure 4 :** *Partie supérieure de l'arbre de classification dans l'espace des dix premiers facteurs*

01\$-----|**-1**-----+  
29\$--+ | |  
24\$--\*+ | |  
27\$--+| | |  
16\$--\*\*|-**2**+ | |  
02\$-----|**-3**\*-----\*-----+  
23\$--+ | | |  
04\$--\*--|**-4**-----\*-----+  
20\$--+ | | |  
21\$--+| | | |  
05\$--\*\*|-**5**-----\*-----+  
30\$--+ | | |  
18\$--\* | | |  
06\$--\*--|**-6**-----+ | |  
19\$-----+| | | |  
10\$--\*--|**-7**-----\*-----\*-----+  
+  
09\$-----|**-8**-----+  
\*-----+  
17\$--+ | | |  
| | | |  
12\$--\*--|**-9**--+ | | |  
| | | |  
14\$--+ | | | |  
| | | |  
22\$--\* | | | |  
| | | |  
15\$--\*--|**10**-- | | |  
\*-----\*+  
13\$-----| | | |  
**11**-----+  
\*--+ | | | |  
26\$--+ | | | |  
| | | |  
25\$--\* | | | |  
| | | |  
28\$--\* | | | |  
| | | |  
03\$--\*--| | | |  
**12**-----+  
--\*-----+ | | | |  
07\$-----| | | |  
**13**-----+  
-----\*-----+ | | | |  
08\$-----| | | |  
**14**-----+  
-----\*-----+ | | | |  
11\$ | | | |

La partition en 15 classes constitue la typologie que nous allons décrire. Elle est obtenue après coupure de l'arbre de classification et optimisation par ré-affectation des individus à la classe dont le centre de gravité est le plus proche; on maximise ainsi l'inertie inter-classe (PROC PARTI de SPADN). Sur la base de ce critère, la partition explique 82% de l'inertie totale du nuage dans l'espace des dix premiers facteurs (voir tableau de la figure 4). Il apparaît donc que la démarche permet de prendre en compte la plupart de l'information significative (c'est à dire hors-bruit) fournie par l'enquête.

### *La caractérisation des classes*

Les coordonnées des centres de classes et les valeurs tests associées, ainsi que les individus les plus proches du centre de chaque classe peuvent fournir la base d'une interprétation "indirecte" des classes à partir des facteurs. Mais pour décrire la typologie il est à la fois plus direct, plus précis et plus riche de revenir à la donnée originale recodée. En ré-affectant sa classe d'appartenance à chaque individu de l'échantillon et en calculant les fréquences de séjours moyennes des individus de la classe dans chacune des modalités au cours de chaque période de recodage, on obtient le profil de mobilité de chaque classe résumé dans un tableau directement lisible et que l'on peut traduire graphiquement <sup>2</sup> (voir section suivante). On peut, avec la même technique, obtenir les profils de classes correspondants à chacune des variables longitudinales illustratives, profils comparables entre eux ainsi qu'au profil d'ensemble de l'échantillon. Enfin, on peut éditer la série de tableaux croisés qui mettent en relation la typologie avec chacune des caractéristiques transversales des individus et des ménages retenues comme illustratives; le pouvoir "explicatif" de chaque caractéristique transversale peut être résumé par la statistique du Khi2 associée au tableau et l'on peut, pour structurer le commentaire, repérer les cellules du tableau ayant les contributions les plus fortes au Khi2 global.

A partir de l'ensemble de ce matériel, on peut donc, pour chaque classe, décrire le comportement résidentiel des enquêtés et dégager, s'il y a lieu, leur trajectoire spatiale spécifique. Cette trajectoire est mise en relation avec la succession des événements du cycle de vie que permettent d'appréhender les variables longitudinales illustratives. On signale également les caractéristiques démographiques et socio-économiques qui complètent la "carte d'identité" de la classe. Des hypothèses sur les dépendances et les inter-relations qui structurent la biographie des individus peuvent être formulées et soumises à vérification. Enfin des propositions d'analyse complémentaires visant à approfondir ces résultats descriptifs peuvent être faites.

## **2. LES RÉSULTATS DE L'ANALYSE**

### **2.1 Le schéma d'ensemble de la mobilité spatiale**

Globalement, la typologie en quinze classes est le reflet de deux phénomènes principaux : la mobilité intra-urbaine à l'échelle des arrondissements et la migration provenant de l'extérieur de l'aire métropolitaine - plus précisément l'âge et le lieu d'arrivée des migrants dans l'a.m. Si l'on considère les deux indicateurs synthétiques du tableau 16 (Imobalc -col. 8- et Pctmig -col. 5-), la corrélation avec le rang des classes apparaît clairement. Sur cette base, on peut distinguer quatre groupes.

1. Très stables (7,9% des individus, voir graphiques 1 à 3 en annexe)

Ce groupe, qui est formé des classes dont la mobilité intra-urbaine a été la plus faible, est entièrement localisé hors de Bogotà : stables à Chia (classe 14 : 4,9%), stables à Madrid (classe 13 : 2,2%) et stables à

## 2. Stables (17,8% des individus, voir graphique 4 à 8 en annexe)

Un peu plus mobiles que les individus du groupe précédent, les classes 9, 10, 11 et 12 sont constituées de noyaux de population stable dans divers quartiers de Bogota : Usaquen (classe 10 : 2,65%), Rafael Uribe et Antonio Nariño (classe 11 : 3,8%), La Candelaria et Santa Fé (classe 12 : 5,9%), auxquels se joignent la classe 2 (1,44%), composée de jeunes migrants stabilisés à Madrid et la classe 9, également jeune, stabilisée à Usaquen. Dans ce groupe, le nombre moyen de changement d'arrondissement de résidence depuis l'entrée dans l'a.m. varie autour de 1.

## 3. Mobilité intra-urbaine moyenne (44,5% des individus, voir graphiques 9 à 13 en annexe)

Le groupe le plus important de la typologie est formé par des individus ayant connu autour de deux changements d'arrondissement depuis leur entrée dans l'a.m. et que caractérisent donc des longs séjours dans certains quartiers, sans pour autant que les résidences à la date de l'enquête soient toujours très concentrées. Ce groupe contient dans l'ordre de la mobilité croissante les classes suivantes :

- Longs séjours dans "autres municipios de l'a.m." (classe 8 : 2.6%)
- Longs séjours à Candelaria et Teusaquillo (classe 7 : 1,9%)
- Longs séjours à Rafael Uribe (classe 5 : 9.1%)
- Longs séjours dans la périphérie Ouest (classe 4 : 8.6%)
- Vieux migrants arrivés entre 20 et 40 ans dans tous les quartiers (classe 1 : 22,3%)

## 4. Mobilité intra-urbaine forte (29,9% des individus, voir graphiques 14 et 15)

Le groupe le plus mobile comprend la classe 6 (7,3%), composée de vieux migrants arrivés jeunes et de natifs qui ont tous connu une forte mobilité dans les quartiers du péri-centre Nord et de la périphérie Nord et Ouest; et enfin la classe 3 (22.6%) qui regroupe les jeunes migrants arrivés entre 10 et 20 ans, dont les itinéraires se jouent souvent dans la périphérie Ouest et Sud. Dans la classe 3, les résidences à la date de l'enquête sont assez concentrées : Bosa, Soacha, Gustavo Restrepo.

Tableau 16 : Indicateurs de mobilité moyenne pour les quinze classes de la typologie

(1) CLASSES	(2) NBINDIV	(3) POIDS	(4) PCTPOID S	(5) PCTMIG	(6) DURMOY	(7) IMOBRES	(8) IMOBAL C
1	217	36156	22,32	95,9	21,3	12,7	7,86
2	25	2326	1,44	69,0	21,8	9,32	4,02
3	196	36557	22,57	83,3	15,8	18,5	14,23
4	72	13890	8,58	39,9	27,6	9,57	7,08
5	65	14706	9,08	30,1	30,0	9,33	6,67
6	89	11889	7,34	54,7	38,6	14,0	11,05
7	32	3039	1,88	12,2	47,4	8,69	6,13
8	34	4244	2,62	24,8	40,8	9,16	6,00
9	37	6515	4,02	40,3	23,6	10,2	5,56
10	25	4294	2,65	11,3	34,8	5,91	2,83
11	26	6157	3,80	25,8	35,6	4,98	3,15

- (1) : Rang de la classe dans la typologie
- (2) : Nombre d'individus de l'échantillon biographique
- (3) : Population extrapolé
- (4) : Pourcentage de la population totale extrapolée
- (5) : Pourcentage d'individus nés hors de l'a.m.
- (6) : Durée moyenne de résidence dans l'a.m.
- (7) : Fréquence des changements de résidence dans l'a.m. (100 x nb. de chgt./nb. d'années de résid. dans l'a.m.)
- (8) : Fréquence des changements d'arrondissement dans l'a.m. (100 x nb. de chgt./nb. d'années de résid. dans l'a.m.)

## 2.2 L'insertion résidentielle et professionnelle des jeunes migrants

### 2.2.1. Les modalités d'insertion résidentielle à l'arrivée à Bogota

Dans la classe 3, la fréquence des statuts "dépendants" au sein du ménage (autres parents du c.m. et autres non parents du c.m.) entre 15 et 24 ans témoigne de l'importance de l'hébergement pour des jeunes migrants lors de l'arrivée à Bogota.

Le fait marquant dans le profil longitudinal de la relation de parenté au chef de ménage est la sortie très rapide du ménage parental (graphique 103), et surtout le fait que ces sorties ne correspondent pas toujours à des entrées directes dans les statuts de c.m. ou de conjoint du c.m. : entre 15 et 24 ans on note des fréquences très importantes dans les statuts d'autres parents et d'autres non parents du c.m. (23% entre 15 et 19 ans, 24,3% entre 20 et 24 ans vs 13,5% et 12,2% aux mêmes âges dans l'ensemble de l'échantillon). De plus le "roulement" est important dans ces statuts de transition : en fait 53,6% des individus de la classe ont connu un épisode de ce type à la date de l'enquête, le fait étant nettement plus fréquent chez les femmes (65%) que chez les hommes (37%). A ce niveau de fréquence, le phénomène apparaît comme une des modalités principales de l'insertion résidentielle des jeunes migrants à leur arrivée à Bogota.

Le profil de la composition des ménages dans lesquels ont vécu les enquêtés résume bien la spécificité du cycle de vie des jeunes migrants de la classe (graphique 106 à comparer au profil d'ensemble de l'échantillon). Aux âges correspondants à l'arrivée des individus dans l'A.M. (entre 10 et 35 ans) c'est la fréquence des "autres types" de ménages qui frappe; cette modalité regroupe les ménages unipersonnels et les ménages sans noyau familial au niveau du c.m. Là encore la fréquence réelle des épisodes de ce type est supérieure à ce que fait apparaître l'observation longitudinale : 72% des individus de la classe sont concernés avec cette fois une fréquence supérieure chez les hommes (84% contre 64% des femmes). Si on rapproche ce phénomène de la fréquence des épisodes dans les statuts d'autres parents ou autres non parents du c.m. (et en retournant à la donnée individuelle pour confirmer l'hypothèse), il apparaît que pour les femmes la quasi totalité des épisodes de vie en ménages sans noyau familial ont lieu sous ces statuts "dépendants" tandis que pour les hommes une fraction importante correspond à un statut de chef de ménage. Ainsi pour 30,4% des hommes de la classe l'accès au statut de c.m. se fait dans un ménage "non familial" (généralement unipersonnel) contre 6,6% seulement pour les femmes.

Le schéma d'insertion résidentiel entre 15 et 30 ans des jeunes migrants apparaît donc d'une part très spécifique par rapport à celui des autres individus de l'échantillon au même âge (marqué par l'hébergement et la vie dans des ménages non familiaux) et d'autre part très différencié selon le sexe (hébergement très fréquent pour les femmes et épisodes en ménages unipersonnel et non familiaux presque systématique pour les hommes).

### 2.2 La mobilité professionnelle particulière des jeunes migrants

Les jeunes migrants ont également des modalités d'insertion professionnelle spécifiques (graphique 109), en particulier en ce qui concerne la mobilité entre emplois salariés et emplois de



on est donc en présence d'un fait statistique d'importance majeure. En revenant aux données individuelles on constate que son explication réside bien dans la mobilité professionnelle vers la catégorie des petits producteurs indépendants et on peut en démontrer précisément le mécanisme.

Sur l'ensemble des individus ayant exercé un emploi à la date de l'enquête (84% de la classe), 35,3% ont occupé au moins un emploi de petit producteur, mais leurs itinéraires professionnels sont de deux types différents. Pour 22,1% d'entre eux, l'activité de petit producteur a été la première activité et dans ce cas, tous l'ont débuté avant 28 ans et 90% avant 25 ans. Les autres (13,2%) y accèdent après d'autres emplois et cette transition s'effectue pour 21% d'entre eux avant 22 ans mais pour 74% entre 22 et 25 ans (5% seulement après 25 ans), ce qui cause la modification brutale du profil d'activité de la classe à cet âge. Notons pour finir que la transition inverse (petits producteurs vers autres emplois) n'est pas pour autant moins fréquente, au contraire : 14,7% des individus ayant eu un emploi à la date de l'enquête l'ont connue, mais tous avant 22 ans; elle contribue donc également à expliquer le profil longitudinal en limitant les densités de présence dans l'état de petit producteur avant 24 ans. Ainsi, ce n'est pas le solde de la mobilité entre les emplois de petits producteurs et les autres qui explique la recomposition observée (au sein de la classe, ce solde est au contraire plutôt favorable aux autres emplois). Le phénomène est dû au décalage dans le temps des deux types de mobilité : on sort des emplois de petits producteurs vers d'autres emplois entre 15 et 22 ans alors que, dans la majorité des cas, on y entre en venant d'autres emplois entre 22 et 25 ans. Bien sur cette catégorie des petits producteurs est très hétérogène et on peut penser que le type d'activité exercée et le revenu ne sont pas les mêmes dans les deux cas...

## CONCLUSION

Aux plans théorique et pratique, lorsqu'on envisage son application à des données biographiques complexes, l'analyse harmonique qualitative satisfait une série de conditions importantes.

Tout d'abord la discrétisation et synchronisation des temps biographiques individuels est modulable et permet diverses stratégies de recodification des données adaptée à différentes problématiques d'analyse de l'enquête. On peut comme on l'a vu adopter d'autres logiques de découpage de la période d'analyse que celle présentée ici. Par exemple, celle du temps calendaire où les bornes des intervalles de recodification sont choisies en fonction de dates qu'on estime déterminantes dans le contexte conjoncturel, social et politique etc. de la période; l'analyse prend alors un caractère plus historique qui fournit un autre point de vue intéressant sur les données (voir par exemple Barbary [1993], pp. 30-64). Une troisième possibilité consiste à adopter un temps biographique plus collectif découpé en fonction d'événements clefs du cycle de vie : les carrières individuelles regroupées dans les classes de la typologie ne sont plus synchronisées exactement selon l'âge des individus mais selon le temps passé depuis (ou avant) un événement donné de la biographie : première arrivée des migrants dans la zone d'enquête, première sortie du domicile des parents, premier accès à la propriété du logement etc. Une analyse complémentaire adoptant cette stratégie seraient très certainement instructive dans le cadre problématique qui nous intéresse ici.

D'autre part, la question du découpage optimal correspondant à chacune de ces logiques, afin de conserver au mieux l'information asynchrone collectée dans l'enquête, peut être posée en terme théorique : optimisation, au sens d'un certain critère statistique, du nombre et des limites des intervalles de codification du processus. Ceci ouvre d'ailleurs le champs à des recherches futures. En s'en tenant au point de vue empirique que nous avons adopté dans ce travail, on a vu que rien n'oblige à choisir des périodes de recodage de durée constante ou une métrique uniforme sur le temps. Au contraire ce sont là deux paramètres importants qui contribuent à la souplesse de la méthode en permettant de l'adapter à la structure particulière des données et à la problématique de l'analyse.

Enfin, l'usage des variables longitudinales illustratives permet tous les regards croisés possibles entre les différents types d'événements biographiques présents dans le questionnaire. Si l'on collecte de plus en plus dans une même enquête plusieurs types de biographies (résidentielles, professionnelles, familiales etc.), c'est évidemment que l'on présume des relations étroites entre elles; on attend alors de la méthode d'analyse qu'elle puisse les mettre en évidence et les décrire précisément. Pour ce faire, l'analyse harmonique qualitative offre deux possibilités.

Une première voie, que nous n'avons pas adoptée ici, consiste à analyser la variable d'état complexe qui résulte du croisement des différentes situations résidentielles, professionnelles, familiales etc. Elle paraît séduisante puisqu'elle donne à tous les chapitres de la biographie le statut de variable active et donc un poids équivalent dans le résultat typologique. Il ne faut pourtant pas croire qu'on fera un bon usage de la méthode en croisant à tout-va les informations biographiques de l'enquête, laissant au calcul algébrique le soin d'exhiber une typologie "complète" qui rende compte de toutes les relations structurant les données. Il y a des limites de

d'une méthodologie récente, dans un domaine où l'accumulation d'expérience est fondamentale, on manque encore beaucoup de pratique dans l'interprétation des résultats de ce type.

La technique des variables illustratives nous semble mieux convenir à l'analyse descriptive conjointe des différents itinéraires. D'un côté elle affirme un a priori problématique - la priorité donnée à la variable active - qui est l'expression d'une démarche expérimentale structurée. De l'autre elle permet de maintenir des normes de représentativité statistique acceptables pour des échantillons qui ne peuvent pas, sauf exception, être très grands (la collecte biographique est complexe et coûte cher!). Outre les éléments purement descriptifs qu'elle permet de dégager, la démarche que nous venons de présenter nous semble également avoir de l'intérêt dans une perspective ultérieure de modélisation. En effet les régularités quasi fonctionnelles observées sur les graphes, dès que les classes atteignent un effectif suffisant, suggèrent très fortement des hypothèses quant aux formes et aux paramètres des fonctions de séjour dans les états biographiques. Ces hypothèses, dont la justification fait souvent défaut dans les démarches de statistique inférentielle, pourront être employées dans la construction de modèles paramétriques ou semi-paramétriques pour estimer directement l'effet de certaines variables exogènes.

En conclusion, si cette expérience d'application parvient à montrer l'intérêt et la grande richesse des données biographiques collectées pour décrire et analyser la mobilité spatiale des citoyens, et mettre à jour les relations qu'elle entretient avec les trajectoires familiales et socio-professionnelles, le mérite en revient bien sûr à la qualité des données, mais également pour une bonne part à la méthode d'analyse. Nous sommes convaincus qu'elle constitue actuellement l'approche statistique la mieux adaptée à la phase descriptive de l'exploitation de ce type d'enquête et elle nous semble le préalable indispensable de toute démarche modélisatrice à l'ambition plus "explicative".



## **Annexes**

Tableau 9 : Fréquences selon l'arrondissement de résidence, la relation de parenté avec le chef de ménage, du statut matrimonial et le nombre d'enfants co-résidant avec l'individu - (situation en 1993 des individus de l'échantillon biographique)

RESIALC			PARENJH			ESTCIV			NUMHIJ		
%	Nro		%	Nro		%	Nro		%	Nro	
col.	obs		col.	obs		col.	obs		col.	obs	
situación en 1993			situación en 1993			situación en 1993					
Usaquen			Jefe de hog			Soltero(a)			268 1		
20.5	202		12.3	112		42.5	439		25.3	268	1
Chapinero			Conyug jh			Unido(a)			2 hijos		
23.0	226		5.4	83		32.8	325		60.7	618	2
Santa fe			Hijo(a) jh			Separado(a)			3 hijos		
13.0	145		9.9	168		16.8	174		9.8	97	3
Bosa			Otro Par.			Viudo(a)			4 hijos		
4.4	43		15.9	87		5.6	59		4.3	48	4
Kennedy			Otro no Par.			Total			5 hijos		
1.7	17		3.3	10		2.3	34		100.0	1031	5
Fontibón			Total						6 hijos		
			2.7	29		100.0	1031				
y mas									Sin hijo		
1.1	13		6.7	66							
Engativa									Total		
36.4	385		0.1	1							
Barrios Uni									6.8		
100.0	1031										
Antonio Nariño											
43											
La Candelaria											
			1.8	24							
Rafael Uribe											
			11.1	53							
Chia											
			8.7	96							
Tabio											
			1.0	53							
Madrid											
			5.6	103							
Soacha											
			8.6	100							
Otr Mun AM											
			0.0	1							
Fuera AM											
			0.2	2							
Total											
			100.0	1031							

Tableau 10 : Fréquences selon le statut d'occupation du logement, le niveau d'étude en cours, la catégorie socio-professionnelle et la composition du ménage - (situation en 1993 des individus de l'échantillon biographique)

OCUPVIV		%	Nro	NIVEDU		%	Nro	CSO		%	Nro	COMPHOG	
col.	obs		col.	obs		col.	obs	col.	obs		col.	obs	
situación en 1993			situación en 1993			situación en 1993			situación en 1993			situación en 1993	
Prop miemb h	65.2	61.4	594	No estudia	88.9	917	No trabaja	43.1	417	Nuclear Comp			
Arriendo	15.2	31.5	336	Preesc o Prim	0.1	1	Director	0.5	10	Nuclear			
Incomp	5.9	6.3	89	Secundario	2.8	27	Prof Especific	2.1	26	Extend comp			
Usufructo	0.7	0.8	12	Universitario	6.1	65	Prop Peq Emp	1.6	19	Extend Incomp			
Otro tipo	13.0	100.0	1031	Otro Tipo	2.0	21	Peq Produc Ind	18.9	181	Otro tipo			
Total	100.0			Total	100.0	1031	Cuadr y Tecn	4.6	54	Total			
								Empleados	5.0				
								Obreros Calif	17.3	186			
								Obreros no Calif	2.8	35			
								Empl. Dom.	4.0	44			
								No especific	0.1	1			
								Total	100.0	1031			

Tableau 11 : Fréquences selon le sexe, la classe d'âge, le statut migratoire et le lieu de naissance - (individus de l'échantillon biographique)





Tableau 12 : Fréquences selon la municipalité de résidence antérieure, la durée de résidence dans le logement de l'enquête, la durée de résidence dans la municipalité de l'enquête et le système de résidence - (individus de l'échantillon biographique)

MUNANTR1			DURVIVIR			DURMUNIR			SISRESIR		
%	Nro		%	Nro		%	Nro		%	Nro	
col.	obs		col.	obs		col.	obs		col.	obs	
Mun. de resid ant.			Durac. en vivi			Durac en munic			Sistema de resid		
Siempre mismo muni.	27.4	303	Menos de 1 ano	21.3	212	Menos de 1 ano	6.2	60	Sist. unipolar		
90.7	938										
Bogota	11.1	151	1-4 anos	33.0	347	1-4 anos	15.6	175	Sis. itinerante		
2.0	16										
Otro municip. A.M.	6.7	82	5-9 anos	15.2	147	5-9 anos	10.2	111	Sis. bipolar. 8-25%		
otra viv.	4.4	52									
Antioquia	1.6	17	10-19 anos	16.1	173	10-19 anos	20.5	194	Sis. bipolar. 25-		
50% otra viv.	2.0	18									
Boyaca	11.6	92	20 anos o mas	14.4	152	20 anos o mas	47.5	491	Sis. bipolar. 50-		
75% otra viv.	0.4	3									
Cundin. fuera A.M.	11.9	109	Total	100.0	1031	Total	100.0	1031	Sis. bipolar. 75-		
92% otra viv.	0.5	4									
Huila		2.6	13								
100.0	1031										
Meta							1.5			24	
Santander	4.9	43									
Tolima	5.0	46									
Valle del Cauca	2.1	21									
Otros dep. Colombia	10.4	98									
Otro pais	3.1	31									
No informa	0.2	1									
Total	100.0	1031									

Tableau 13 : Fréquences selon le type de logement occupé, le nombre de pièces utilisées pour dormir, le nombre de personnes du ménage et l'indicateur de promiscuité dans le logement- (ménages de l'échantillon biographique)

<b>TIPOVIV</b>	% col	Obs	<b>NUMDORM</b>	% col	Obs	<b>PERSHOG</b>	% col	Obs	<b>HACINTOR</b>
Tipo de viv.			Num de dorm			Tamaño hog			Hacinamiento
Casa independiente	50.8	578	1 cuarto	29.9	314	1 persona	4.5	52	Menos de 1
persona	2.8	24							
Apartamento	31.7	308	2 cuartos	37.7	374	2 personas	11.0	115	1 a 1.9
personas	38.6	406							
Cuarto	11.2	105	3 cuartos	19.8	213	3 personas	19.9	188	2 a 2.9
personas	28.4	307							
Viv. ed. no habitable	0.5	3	4 cuartos	9.5	93	4 personas	23.1	246	3 a 4.9
personas	24.5	225							
Edif. en construccion	4.8	28	5 cuartos o mas	3.1	36	5 personas	20.5	221	5 personas o
mas	5.6	69							
Otro	1.1	9	No informa	0.0	1	6 personas	10.5	103	Total
100.0	1031								
Total		100.0	1031	Total		100.0	1031	7 personas	5.3
53									
						8 personas	2.4	28	
						9 personas o mas	2.8	25	
						Total	100.0	1031	

Tableau 14 : Fréquences selon le sexe du c.m., la classe d'âge du c.m., le statut migratoire du c.m. et l'âge moyen des enfants du c.m. - (ménages de l'échantillon biographique)

SEXOCM	% col	Obs	EDADCM	% col	Obs	STMIGCM	% col	Obs	EDADHIJR
Hombre	80.3	805	5-19 años	0.5	7	Siemp.misma vivienda	1.9	21	0-5 años
Mujer	19.7	226	20-29 años	17.2	169	Siemp.mismo municipio	19.4	217	5-10 años
Total	100.0	1031	30-39 años	26.3	258	Nac:muni.,ant:en A.M.	1.4	20	10-15 años
			40-59 años	37.1	396	Nac:muni.,ant:fuera A.M.	4.5	61	15-20 años
			60 años o más	18.9	201	Nac:en A.M.,ant:en A.M.	6.2	72	20-30 años
			Total	100.0	1031	Nac:en A.M.,ant:fuera A.M.	0.8	12	30-40 años
						Nac:fuera A.M.,ant:en A.M.	10.0	142	40 años o mas
						Nac:fuer A.M.,ant:fue A.M.	55.3	483	Sin hijos en hogar
						No informa	0.4	3	Total
						Total			100.0

Tableau 15 : Fréquences selon le statut d'occupation du logement - (ménages de l'échantillon biographique) et le mode d'acquisition du logement - (ménages propriétaires de l'échantillon)

<b>TENVIVR</b>	% col	Obs	<b>ADQUIVIR</b>	% col	Nro
Tenencia de viv.			Adquis viv.		
Propiedad tot. pagada	55.0	534	Herencia	9.5	71
Propiedad pagando	7.4	67	Herencia + otro rec.	0.8	7
Arriendo (con contr.)	11.2	124	Prestamo	10.5	78
Arr. (sin cont -par.)	7.1	67	Prestamo + rec. prop.	9.7	70
Arr. (sin cont. -no par)	12.3	137	Recursos propios	65.8	353
Arr. (no inf. cont.)	0.7	9	Otra forma	2.6	16
Usufructo (pariente)	3.5	42	No informa	1.2	6
Usufructo (no pariente)	2.3	45	Total	100.0	601
Otro	0.4	5			
No informa	0.0	1			
Total	100.0	1031			

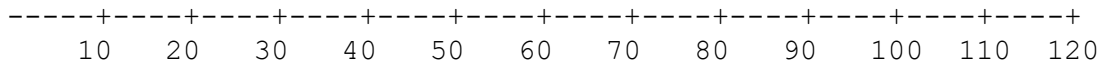
**Figure 1 :** Distribution des changements d'états de la variable RESIALC selon l'âge des individus et découpage temporel adopté pour l'analyse de la variable

RESIALC			
Cum.		Cum.	Freq
Midpoint			
Freq	Percent	Percent	
0	****		8
8	0.28	0.28	
1	*****		53
61	1.86	2.14	
2	*****		44
105	1.54	3.68	<b>1 :0-5 ans</b>
3	*****		39
144	1.37	5.05	
4	*****		33
177	1.16	6.21	
5	*****		50
227	1.75	7.96	
6	*****		44
271	1.54	9.50	
7	*****		49
320	1.72	11.22	<b>2 :6-9 ans</b>
8	*****		53
373	1.86	13.08	
9	*****		58
431	2.03	15.11	
10	*****		69
500	2.42	17.53	
11	*****		42
542	1.47	19.00	<b>3 :10-12 ans</b>
12	*****		68
610	2.38	21.39	
13	*****		62
672	2.17	23.56	<b>4 :13-14 ans</b>
14	*****		77
749	2.70	26.26	
15	*****		102
851	3.58	29.84	<b>5 :15-16 ans</b>
16	*****		96
947	3.37	33.20	
17	*****		117
1064	4.10	37.31	<b>6 :17-18 ans</b>

23	*****	88
1730	3.09 60.66__	
24	*****	93
1823	3.26 63.92 <b>10 :24-25 ans</b>	
25	*****	83
1906	2.91 66.83__	
26	*****	86
1992	3.02 69.85	
27	*****	76
2068	2.66 72.51 <b>11 :26-28 ans</b>	
28	*****	52
2120	1.82 74.33__	
29	*****	54
2174	1.89 76.23	
30	*****	79
2253	2.77 79.00 <b>12 :29-31 ans</b>	
31	*****	42
2295	1.47 80.47__	
32	*****	47
2342	1.65 82.12	
33	*****	44
2386	1.54 83.66 <b>13 :32-35 ans</b>	
34	*****	45
2431	1.58 85.24	
35	*****	52
2483	1.82 87.06__	
36	*****	39
2522	1.37 88.43	
37	*****	27
2549	0.95 89.38	
38	*****	30
2579	1.05 90.43	
39	*****	25
2604	0.88 91.30 <b>14 :36-42 ans</b>	
40	*****	32
2636	1.12 92.43	
41	*****	18
2654	0.63 93.06	
42	*****	16
2670	0.56 93.62__	
43	*****	18
2688	0.63 94.25	
44	*****	22



49	*****		17
2774	0.60	97.27	
50	*****		10
2784	0.35	97.62	
51	*****		9
2793	0.32	97.93	
52	*		2
2795	0.07	98.00	
53	**		4
2799	0.14	98.14	
54	***		5
2804	0.18	98.32	
55	***		5
2809	0.18	98.49	
56	**		4
2813	0.14	98.63	
57	**		3
2816	0.11	98.74	
58	**		4
2820	0.14	98.88	
59	****		7
2827	0.25	99.12	
60	***		5
2832	0.18	99.30	
61	**		4
2836	0.14	99.44	
62	*		2
2838	0.07	99.51	
63	***		5
2843	0.18	99.68	
64	*		1
2844	0.04	99.72	
65	**		3
2847	0.11	99.82	

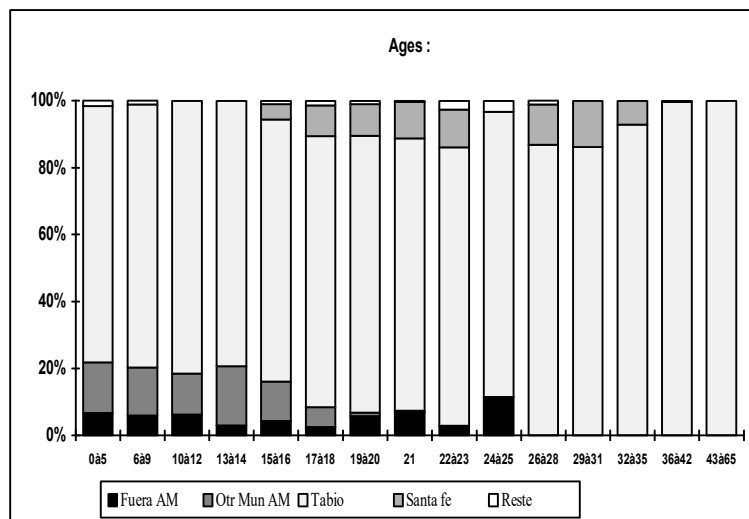


Frequency

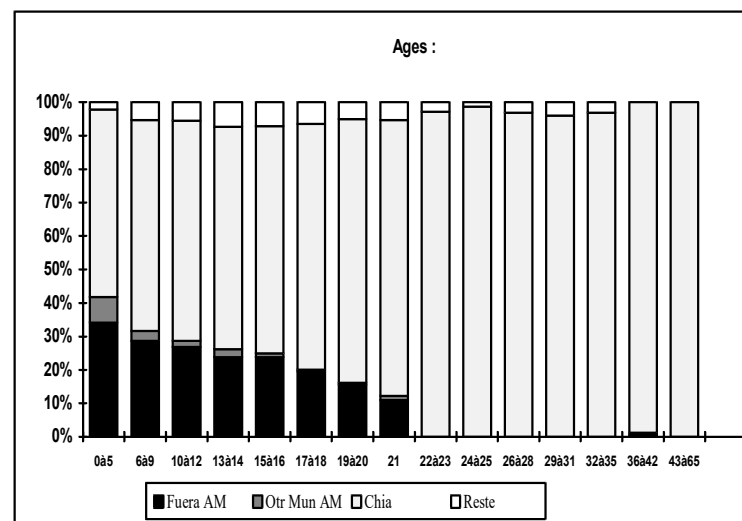


## Les groupes stables et stabilisés dans les municipalités périphériques (Tabio, Chia, Madrid)

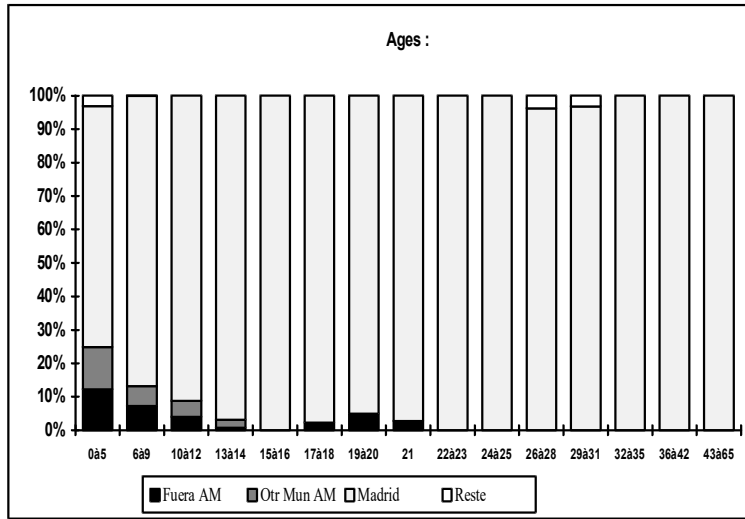
Graphique 1 : Classe 15, stables à Tabio



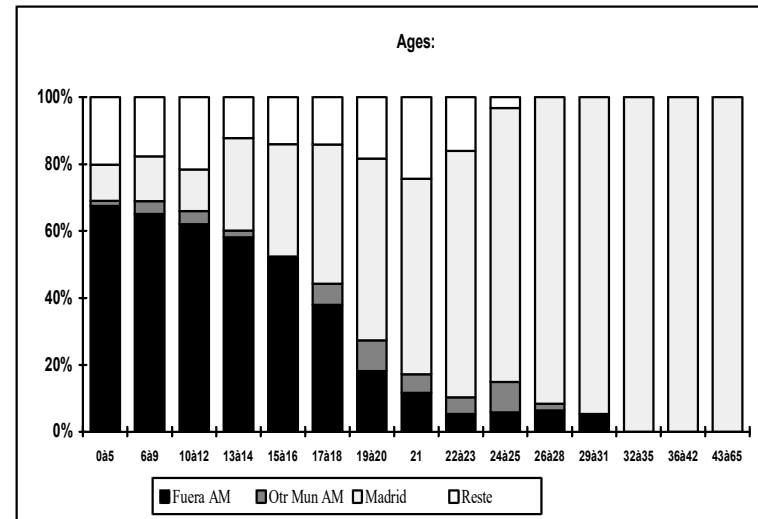
Graphique 2 : Classe 14, stables à Chia



**Graphique 3 : Classe 13, stables à Madrid**

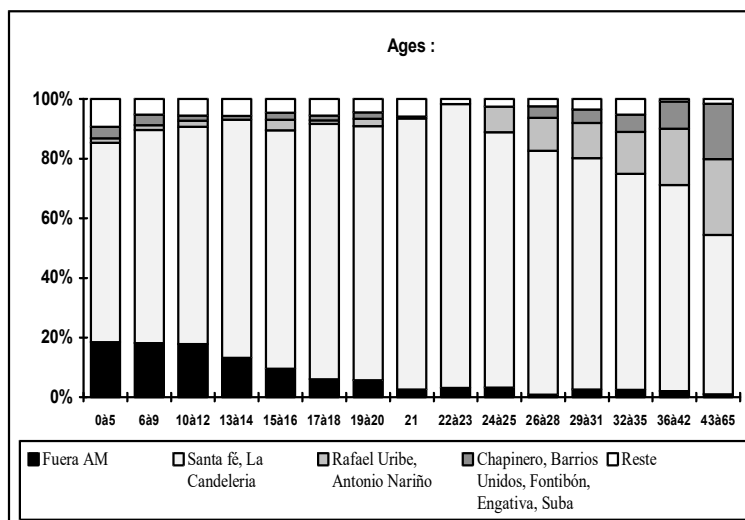


**Graphique 4 : Classe 2, stabilisés à Madrid**

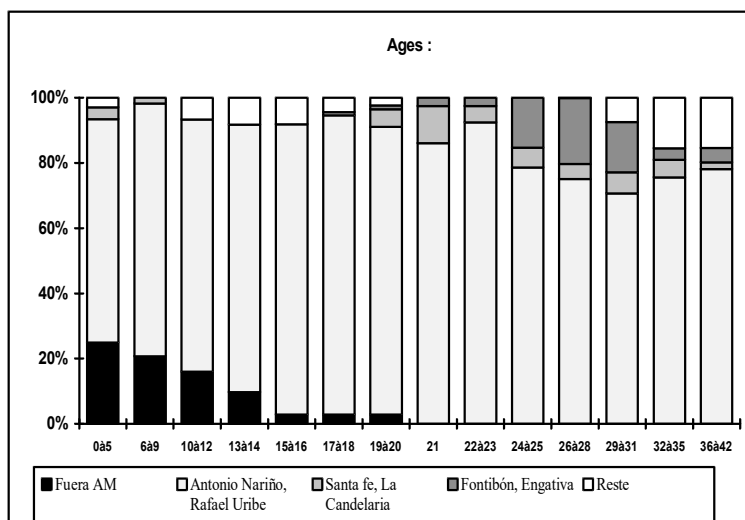


## Les groupes stables et stabilisés dans Bogota

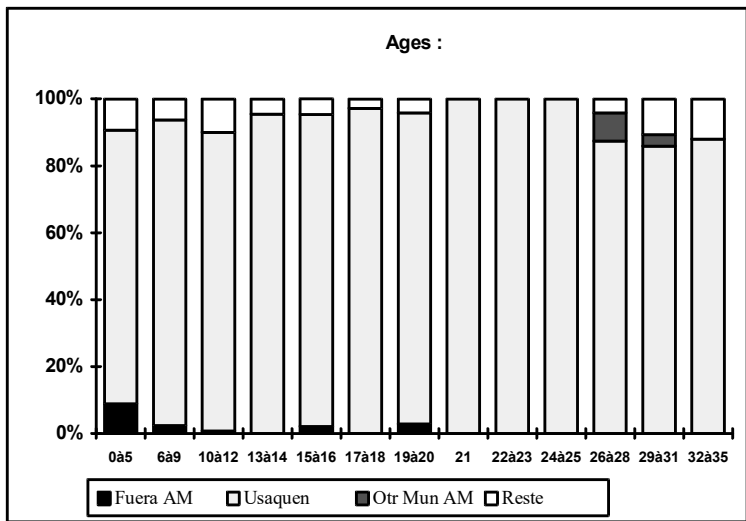
**Graphique 5 : Classe 12, stables à Santa Fé et La Candelaria**



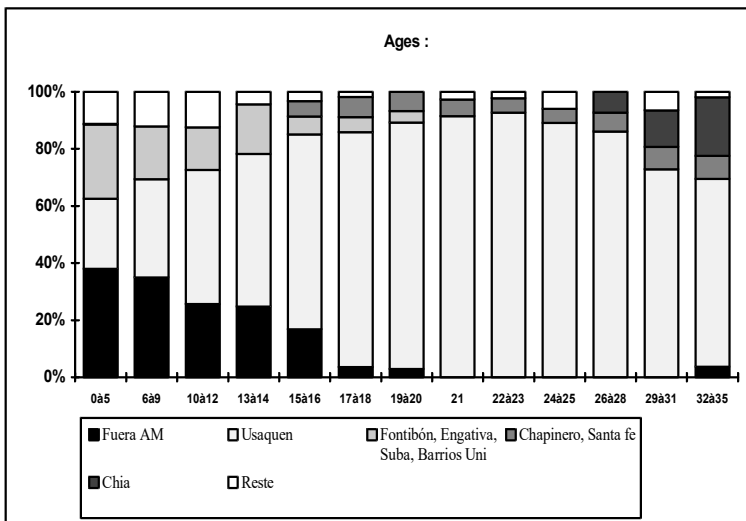
**Graphique 6 : Classe 11, stables à Antonio Nariño et Rafael Uribe**



**Graphique 7 : Classe 10, stables à Usaquen**

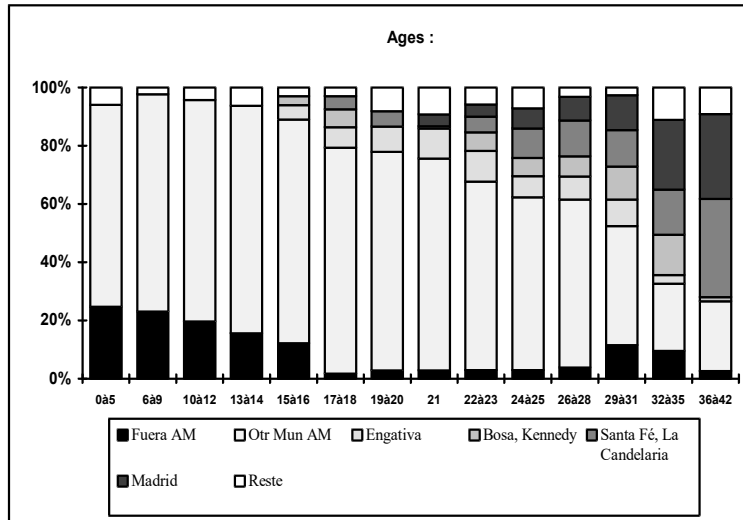


**Graphique 8 : Classe 9, stabilisé à Usaquen**

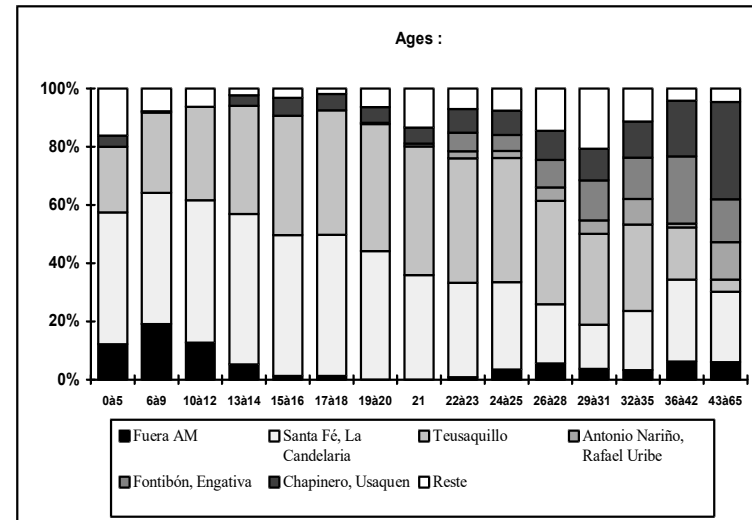


**Les groupes à mobilité intra-urbaine moyenne**

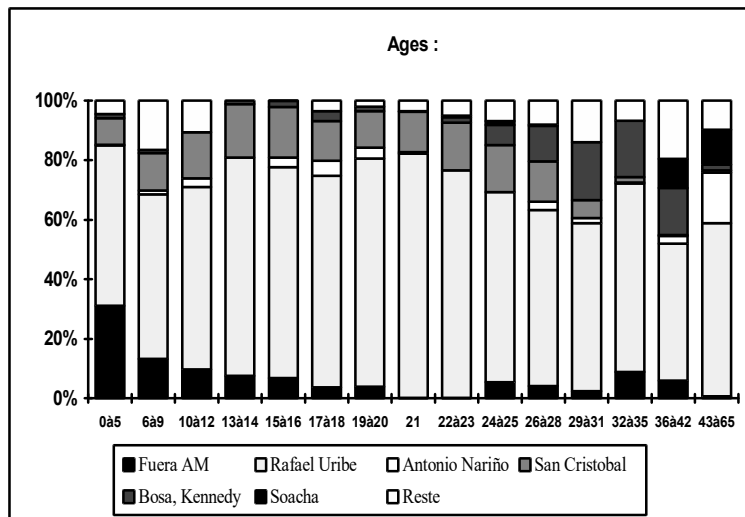
**Graphique 9 : Classe 8, longs séjours d'autres municipios de l'a.m.**



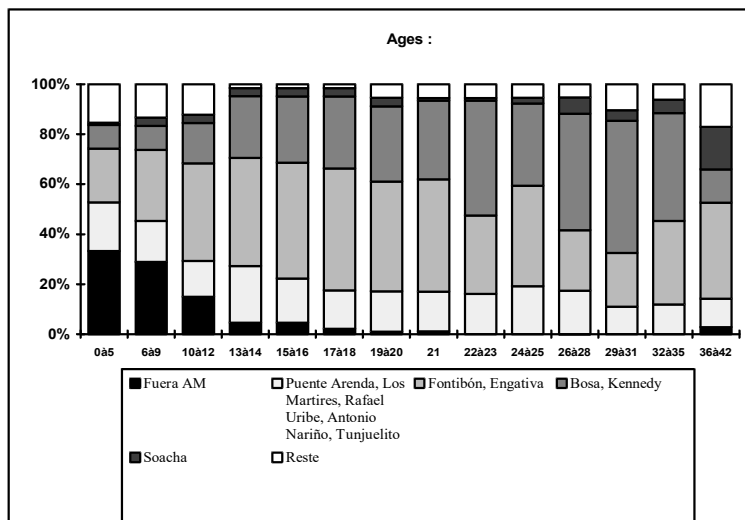
**Graphique 10 : Classe 7, Longs séjours à La Candelaria et Teusaquillo**



**Graphique 11 : Classe 5, longs séjours à Rafael Uribe et San Cristobal**



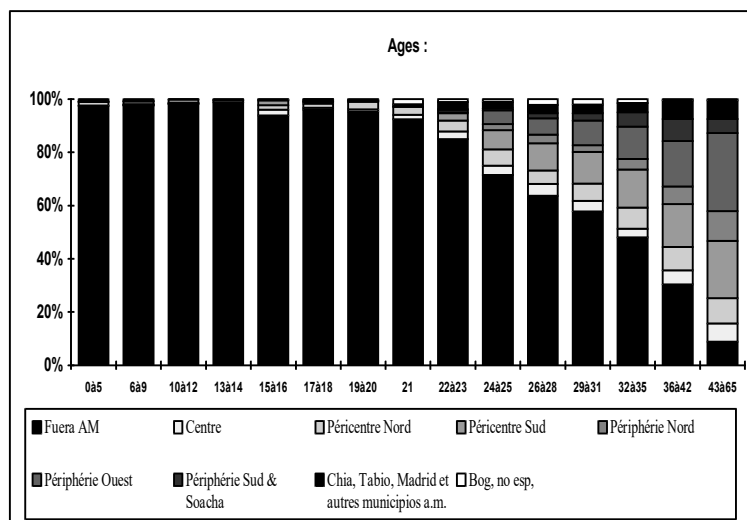
**Graphique 12 : Classe 4, long séjours dans la périphérie Ouest**



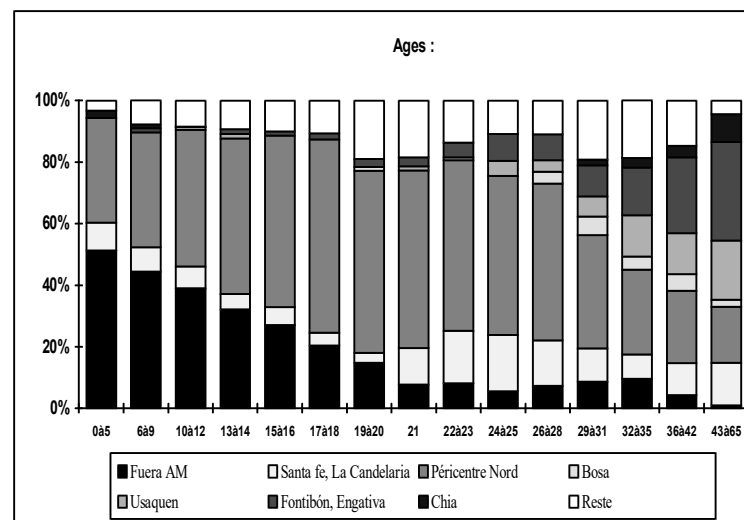


## Les migrants et les groupes à mobilité intra-urbaine forte

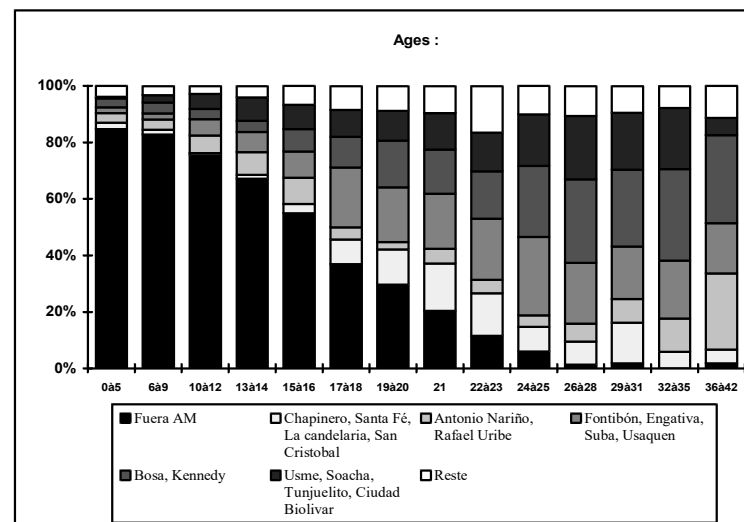
Graphique 13 : Classe 1, vieux migrants arrivés entre 20 et 40 ans



Graphique 14 : Classe 6, vieux migrants arrivés avant 20 ans



Graphique 15 : Classe 3, jeunes migrants arrivés entre 10 et 25 ans





**Notes bibliographiques :**

ARIÑO M., ROSAS M.E., TORRADO S., [1987] : Metodología para el análisis de la estructura socio-ocupacional argentina (1960-1980). Consejo Federal de Inversiones, Documentos de la Secretaria General, Buenos Aires, 31p.

BARBARY O. (1993) avec la collaboration de E.H.A. DIOP, "L'insertion urbaine : le cas de Dakar, rapport d'exploitation d'enquête", Dakar, 213 p. multigr.

BENZECRI J.P. et al. [1973], "L'analyse des données. Tome 2 : L'analyse des correspondance" DUNOD, Paris, 632 p.

BERET P. [1988] : Analyse de données chronologiques relatives à l'insertion professionnelle, in "Les Cahiers de l'Analyse des Données" vol. XIII, n° 2, pp. 159-174.

COURGEAU D., LELIEVRE E. [1989] : Analyse démographique des biographies, éditions de l'INED, Paris, 268 p.

COX D.B. [1972] : Regression models and life tables (with discussion), in "Journal of Royal Statistical Society", B 34, pp. 187-220.

DEGENNE A., LEBEAUX M.O., MOUNIER L. [1995] : Construction d'une typologie de trajectoires à partir de l'enquête de suivi des jeunes des niveaux V, Vbis et VI. Communication aux Journées CEREQ-LASMAS-IDL sur les données longitudinales dans l'analyse du marché du travail, Caen, 28-29 juin 1995, 14 p.

DEVILLE J.C., SAPORTA G. [1980] : Analyse harmonique qualitative, in "Data Analysis and Informatics", E. DIDAY et al. éditeurs, North Holland Publishing Compagny, pp. 375-389.

DEVILLE J.C. [1982] : Analyse des données chronologiques qualitatives, comment analyser les calendriers ? "Annales de l'INSEE", n° 45, pp 45-104.

DROESBEKE J.J., FICHET B., TASSI Ph. éditeurs [1989] : Analyse statistique des durées de vie, modelisation des données censurées, ECONOMICA, Paris, 282 p.

DUREAU F., BARBARY O., FLOREZ C.E. [1994] : La movilidad de las poblaciones y su impacto sobre la dinámica del área metropolitana de Bogotá, resultados preliminares de la encuesta cuantitativa. Documento de trabajo No 3, CEDE/ORSTOM, Bogotá, 309 p.

FLOREZ C.E, MENDEZ R., [1994] : Boletín de estadísticas sobre hogar y familia en Colombia, 1972-1992. Informe final, Universidad de los Andes, CEDE, Bogota, 43p.

VAN DER HEIJDEN P. G. M. [1987] : Correspondence analysis of longitudinal categorical data, DSWO PRESS, Leiden.