

De l'espace rural à l'espace urbain

Robert CHAPUIS

1973 - Etudes Rurales n°49-50 - Problèmes de typologie

Qu'est-ce que l'espace urbain ? Qu'est-ce que l'espace rural ? Comment définir des types d'espace urbain, des types d'espace rural ? Comment délimiter ces espaces ? Voilà quelques questions auxquelles on a bien souvent essayé de répondre.

On a donné deux sortes de réponses. Le plus souvent on a tenté de définir, d'une façon conceptuelle, la ville, la campagne, les types d'espace urbain, les types d'espace rural. A la suite de Ratzel, les géographes ont proposé de multiples définitions de la ville. Les sociologues se sont également efforcés de définir le phénomène urbain et d'en marquer les limites. Et bien d'autres. Il faut d'ailleurs remarquer que ce que l'on définit généralement, c'est la ville ; même les chercheurs qui s'intéressent à l'espace rural acceptent le plus souvent que celui-ci ne soit qu'un espace relictuel, un espace non urbain, comme s'ils faisaient confiance à d'autres pour leur limiter un territoire qui, il est vrai, leur est moins mesuré.

Avec le déferlement urbain contemporain, avec « la pénétration physique de l'urbanisation en milieu rural, les échanges sans cesse plus nourris et diversifiés entre ville et campagne... surtout dans les sociétés industrielles »¹, on a substitué, à la simple dichotomie ville-campagne, des concepts plus souples, des typologies plus nuancées qui tiennent compte des espaces intermédiaires : aire urbaine, Umland, etc.².

Ces typologies abstraites et non chiffrées ont certes leur intérêt pédagogique et scientifique. Cependant on pourrait dire de ces tentatives ce que G. Chabot dit des essais de définition de la ville « aucune [définition] n'apparaît entièrement satisfaisante au regard de la géographie urbaines ³[...] il reste derrière toute définition un résidu impossible à cerner de façon précise »⁴. D'ailleurs ces définitions abstraites sont d'un maigre secours pour qui veut délimiter dans l'espace et cartographier les types ainsi reconnus.

La deuxième voie pour une recherche de typologies a été plus prosaïque. Plutôt que de définir des types abstraits, synthétiques et universels, on a tenté, dans

1 O. DOLLFUS, L'espace géographique, Paris, KW, 1970, 126 p.

2 J.-B. HARRIER, Citadins et ruraux, Paris, PUE, 1964, 128 p.

3 G. CHABOT, Les villes, Paris, A. Colin, 1948, 224 p.

4 G. CHABOT et J. BEAUJEU-GARNIER, Traité de géographie urbaine, Paris, A. Colin, 1964, 493 p.

un cadre restreint et à l'aide de critères chiffrés, de reconnaître, hic et nunc, des types d'espaces précis et d'en marquer les limites.

C'est cette deuxième voie de recherche qui va nous intéresser ici. Notre but est de donner un aperçu de quelques méthodes utilisées jusqu'à présent, ou maintenant disponibles, pour bâtir une typologie synthétique, à base chiffrée, des espaces ruraux et urbains. Nous insisterons surtout sur l'espace rural et ses limites avec l'espace urbain. Nos exemples seront pris en Europe et même le plus souvent, en France.

En allant du plus simple au plus complexe, nous passerons des typologies fondées sur un seul critère à celles qui se fondent sur de multiples critères et des typologies utilisant des méthodes statistiques simples (que nous appellerons « classiques ») à celles qui mettent en œuvre des outils mathématiques complexes.

Les typologies « classiques » à un seul critère chiffré

Des nécessités administratives, statistiques ou autres ont obligé depuis longtemps à distinguer nettement entre villes et campagne. On a donc cherché très tôt des critères simples qui permettraient cette dichotomie rudimentaire.

Ce sont des critères démographiques qui ont été le plus souvent utilisés, en particulier la taille de l'agglomération. A l'intérieur d'un périmètre communal, dès que la population atteint une certaine masse, on passe de la campagne à la ville. Le critère de taille, encore à la base de la plupart des statistiques nationales, est sujet à de nombreuses critiques. Il suffira de rappeler que le seuil minimal pour qu'il y ait ville varie de 250 pour le Danemark à 10 000 pour l'Espagne et la Grèce, et surtout qu'il n'y a pas de liaison obligatoire entre la population d'une agglomération et son rôle fonctionnel, pourtant essentiel.

On a pensé à la densité de population. En France la densité de la population des villes atteignait, en 1962, 627 hab./km², contre 34 pour les campagnes. Cependant les exceptions sont nombreuses, à grande comme à moyenne échelle. La population considérée comme rurale du département du Nord atteignait à la même date 102 hab./km², c'est-à-dire un chiffre presque équivalent à celui des communes urbaines de moins de 5 000 habitants (145 hab./km²) sans compter les cas extrêmes d'Arles, qui avec 45 000 habitants n'avait que 55 hab./km²⁵. Même si l'on ne tient compte que de la densité de population sur l'espace bâti, il n'est pas évident que certaines banlieues pavillonnaires aient des densités supérieures à celles de certains villages aux maisons bien groupées ; d'ailleurs les immeubles collectifs sont maintenant nombreux à la campagne. La densité de population est donc à elle seule insuffisante pour typer un espace.

D'autres critères ont été essayés. Notre propos n'est pas de les analyser ici. Rappelons seulement que des voies nouvelles sont actuellement prospectées, par

⁵ J.-C. BONTRON et N. MATHIEU, *L'espace rural français*, Paris, SEGESA, 1968, 49 p. et annexes.

exemple grâce à l'étude des finances communales. En 1966, les ministères de l'Intérieur et de l'Agriculture avaient, pour ce qui est des travaux d'alimentation en eau potable, classé en communes rurales celles dont la valeur du centime sur les quatre dernières années était inférieure à 10 francs, en communes urbaines celles pour qui cette valeur était supérieure à 10 francs. On avait ainsi 954 communes urbaines en France contre 2 124 selon les critères habituels de l'INSEE⁶. Plus récemment P. Limousin note que connaissant la valeur et la structure du potentiel fiscal [des communes], il est possible d'en tirer une typologie. La valeur du principal fictif constitue un premier critère de hiérarchisation ; dans la mesure où il exprime la richesse en matière imposable, il permet de fixer le degré d'urbanisation auquel une commune est parvenue ». Dans le département du Val-d'Oise, il distingue cc une zone rurale dans laquelle les communes ont un principal inférieur à 409 francs et une zone urbaine »⁷. Mais il faut remarquer, là encore, que ces critères peuvent difficilement être utilisés seuls. La délimitation des ministères aboutit à restreindre par trop le domaine urbain et P. Limousin montre lui-même que sa typologie souffre un certain nombre d'exceptions.

La recherche d'un seul critère caractéristique présente donc l'avantage de la simplicité et de la rapidité pour la recherche documentaire comme pour la représentation cartographique. Cependant les phénomènes géographiques ne se laissent pas aussi simplement corseter. Aucun critère ne peut, à lui seul, donner la clé d'une typologie synthétique de l'espace.

On a donc senti la nécessité de combiner plusieurs critères pour cerner d'une façon souple les réalités. Ces essais ont d'abord été tentés avec des méthodes statistiques simples. Ce sont ceux que nous analyserons en premier.

Les typologies « classiques » à multiples critères

Les typologies à quelques critères

Un certain nombre d'instituts nationaux de statistiques, qui avaient longtemps appuyé leur typologie sur un ou deux critères, ont éprouvé le besoin d'assouplir leurs définitions. Par souci de simplicité et d'efficacité il leur est cependant, dans la plupart des cas, interdit de multiplier le nombre des critères. Ils ont donc souvent mis au point une typologie à quatre ou cinq critères.

L'INSEE, par exemple, définissait, en 1954, comme urbaines les communes dont la population, agglomérée au chef-lieu, atteignait ou dépassait 2 000 habitants ; depuis 1962 la définition des agglomérations urbaines s'appuie sur quatre critères : l'effectif total de la commune, la densité de sa population, son taux d'accroissement

⁶ CENTRE DE RECHERCHES ÉCONOMIQUES ET SOCIALES APPLIQUÉES, Organisation de l'espace rural, Paris, CRESA, 1966, 46 p.

⁷ P. LIMOUZIN, « Croissance et dynamisme des communes de la banlieue Nord-Ouest de Paris », Analyse de l'Espace, 1972 (1), janv., pp. 16-62.

au cours des dernières années, son taux de population agricole. L'INSEE encore, peu avant le recensement de 1962, a été conduit à définir un type intermédiaire entre agglomération urbaine et campagne profonde : les zones de peuplement industriel et urbain ; dans la définition de ces ZPIU interviennent la composition de la population active, les migrations journalières, la présence d'établissements industriels et la population totale.

En Belgique, étaient rurales les communes dont la densité de population était inférieure à 200 hab./km² ainsi que celles de 200 à 1 000 hab./km² dont la population active de 14 à 65 ans était aux deux tiers agricole. Depuis 1967, ce pays a adopté une classification analogue à celle des Pays-Bas où interviennent quatre critères : physique ou morphologique (implantation compacte dans le paysage), socio-économique (composition de la population active), fonctionnel (établissements ayant une fonction distributive régionale), sociologique et psycho-social (modes de vie et mentalité)⁸.

Pour utiles qu'elles soient dans le cadre d'un découpage pragmatique de l'espace, ces typologies restent rigides et prêtent, dans le détail, à de multiples discussions. On a donc songé, surtout au niveau de la recherche théorique, à prendre en compte non pas seulement quelques critères, mais parfois plusieurs dizaines pour coller au mieux à des réalités complexes. Deux problèmes se posent. Quels critères faut-il sélectionner ? Comment mettre en œuvre ces critères pour aboutir à une typologie ? Nous illustrerons le premier problème par les travaux du Groupe d'étude de l'urbanisation des campagnes⁹, le second par l'étude de R. Mols sur la périphérie de Bruxelles¹⁰.

Les travaux du Groupe d'étude de l'urbanisation des campagnes

Le groupe a travaillé, sous la direction d'E. Juillard, à sélectionner, pour la France, des critères d'urbanisation des campagnes qui devraient permettre de définir une typologie allant des campagnes les plus profondes aux plus urbanisées. Après avoir éliminé certains critères, comme le bilan migratoire, la proportion des jeunes, le pourcentage des navetteurs, les résidences secondaires, l'orientation politique, qui ne lui paraissent avoir qu'un intérêt local, le groupe a retenu provisoirement les critères suivants : indices de consommation, de mentalité, physiologiques, de développement socio-économique et de rapidité d'évolution.

Indices de consommation

⁸ G. HOYOIS, *Sociologie rurale*, Paris, Éditions universitaires, 1958, 453 p.

⁹ GROUPE D'ÉTUDE DE L'URBANISATION DES CAMPAGNES, « Essai de mise au point des indices à utiliser pour une étude comparée en France », 1969, ronéo.

¹⁰ R. MOLS, « La périphérie bruxelloise. Jalons statistiques pour aider à sa délimitation », *Bulletin de la Société belge d'études géographiques*, Bruxelles, 1969, XXXVIII (2), pp. 183-252.

- nombre de kWh basse-tension consommés par 100 habitants d'une commune en une année ;
- pourcentage par commune de ménages ayant une auto ;
- pourcentage par commune de ménages ayant payé une redevance télévision ;
- montant des factures annuelles de téléphone pour 100 ménages d'une commune (ou à défaut par groupement téléphonique) ;
- nombre de salons de coiffure dames ou mixtes pour 1 000 habitants ;
- indices difficilement chiffrables : dépenses pour soins du corps ; consommation de fruits, fromages, produits diététiques, whisky, cigarettes étrangères, etc.

Indices de mentalité

- pourcentage par commune de femmes allant accoucher à l'extérieur ;
- pourcentage par commune des voix de gauche aux élections générales (n'est sans doute valable que dans les régions industrialisées) ;
- montant pour 100 ménages des crédits à la consommation distribués par la CETELEM ;
- nombre absolu par commune de licences pour les catégories de sport suivantes : football, basket, judo, boules, gymnastique, cycle ;
- montant des droits payés par commune à la Société des auteurs-compositeurs et éditeurs de musique ;
- nombre de donneurs de sang pour 100 habitants d'une commune ;
- pourcentage des non-agriculteurs dans les conseils municipaux ;
- nombre brut de centimes additionnels par commune ;
- évolution du nombre de ces centimes.

Indices physiologiques

- pourcentage de logements non agricoles construits depuis 1949, par commune ;
- nombre de lotissements effectués par commune depuis 1962 ;
- pourcentage de logements, par commune, ayant un WC à l'intérieur.
- pourcentage de logements, par commune, ayant une salle-d'eau ;
- part relative, par commune, des permis de construire concernant des commerces par rapport au nombre total de permis.

Indices de développement socio-économique

- densité (hab./km²) de la population par commune ;
- pourcentage de la population active non agricole par commune ;
- pourcentage, par commune, de la population active âgée de moins de 50 ans ;

- pourcentage, par commune, de la population domiciliée auparavant dans une autre commune ;
- pourcentage, par commune, de garçons et de filles poursuivant des études au-delà de la scolarité obligatoire ;
- pourcentage des chefs d'exploitation, par commune, ayant au moins un brevet de technicien agricole ;
- nombre d'usines situées à moins de 8 km de la commune ;
- durée du trajet par autobus ou chemin de fer vers la « ville » la plus proche ; nombre de trajets aller et retour par jour.

Indices de rapidité d'évolution

- évolution, par commune, du nombre de scolarisés au-delà de l'âge obligatoire ;
- accroissement du nombre d'abonnés au téléphone depuis 1962 ;
- accroissement du nombre de vignettes automobiles délivrées depuis 1962.

C'est une première approche qui nécessitera certainement des retouches, ne serait-ce que parce que certains chiffres ne peuvent être obtenus, au niveau de la commune, que par enquête ; or, de tels travaux doivent s'appuyer essentiellement sur la documentation existante. Ces essais donnent cependant une idée de l'extrême diversité des critères utilisables.

Cependant, alors que la sélection du critère était la phase essentielle de la recherche en cas d'utilisation d'un seul critère, au fur et à mesure que l'on intègre des variables plus nombreuses, c'est la mise en œuvre de ces variables qui pose les problèmes les plus difficiles à résoudre. C'est ce que nous allons voir à propos du travail de R. Mols.

Une typologie de la périphérie bruxelloise

R. Mols a, en principe, pour but de fixer la limite entre l'agglomération de Bruxelles et la campagne ; mais, il le dit lui-même, comme « il n'est pas possible de tracer une ligne de démarcation nette et incontestable entre le domaine de la ville et celui du plat pays »¹¹, il recherche, en fait, une typologie de la périphérie qui ira des communes les plus urbaines aux communes les plus rurales. Première étape : en fonction de son expérience personnelle, de sa connaissance du terrain et des possibilités de documentation, l'auteur choisit 42 critères de départ qu'il appliquera à 90 communes de la périphérie bruxelloise. Deuxième étape : il en combine certains pour arriver à vingt critères synthétiques ; il n'est peut-être pas inutile d'en donner la liste à titre documentaire.

¹¹ Toutes les citations de ce sous-chapitre sont de R. MOLS, art. cit.

Quatre critères démographiques :

1. Rythme de croissance de la population.
2. Balance migratoire particulièrement favorable.
3. Intensité de la mobilité migratoire.
4. Niveau élevé et croissance accélérée de la densité démographique.

Quatre critères de provenance :

5. Proportion élevée et croissante de cc natifs-Bruxellois ».
6. Proportion élevée d'habitants dont la résidence précédente était Bruxelles.
7. Proportion importante et croissante d'étrangers de naissance ou de nationalité.
8. Répartition proportionnelle des cc natifs-provinciaux » s'écartant le moins possible de celle constatée à Bruxelles.

Quatre critères professionnels :

9. Pourcentage des navetteurs vers Bruxelles sur la population active.
10. Pourcentage des navetteurs vers Bruxelles sur tous les navetteurs sortants.
11. Fréquence et concentration spatiale des navetteurs venant de Bruxelles travailler dans cette localité.
12. Fréquence et densité peu élevée des travailleurs de la catégorie cc agriculture et pêche ».

Trois critères « instructionnels » :

13. Niveau de formation scolaire de la population « instruite ».
14. Fréquence des migrations scolaires vers Bruxelles.
15. Fréquence des élèves poursuivant leurs études jusqu'aux niveaux les plus élevés.

Trois critères domiciliaires :

16. Forte densité des constructions, en particulier des maisons servant à l'habitation.
17. Accroissement du nombre de maisons d'habitation au cours des vingt dernières années.
18. Fréquence de quelques caractéristiques domiciliaires exprimant une influence urbaine.

Deux critères spéciaux :

19. Facilité des moyens de communication publics avec Bruxelles.
20. Montant du revenu moyen par contribuable.

Le critère 1, par exemple, est effectivement synthétique puisqu'il est constitué par quatre critères de départ, c'est-à-dire par l'augmentation de la population au cours de quatre périodes : 1901-1930, 1931-1947, 1948-1961, 1962-1968. Ces quatre périodes, d'ailleurs, n'interviennent pas de la même façon. Par exemple, les deux plus anciennes ont chacune un coefficient 2, les plus récentes un coefficient 3. Beaucoup de critères de départ sont ainsi pondérés à l'intérieur des critères synthétiques.

Après ce délicat travail on peut passer à la troisième étape. En effet, les valeurs prises par les différents critères ne sont absolument pas comparables puisque certaines sont des pourcentages, d'autres des valeurs absolues. Il faut donc transposer les séries de valeurs brutes obtenues pour chaque critère en autant de séries de valeurs-indices qui pourront se comparer et se totaliser pour aboutir à l'indice global final. Pour chaque critère l'auteur établit donc « une grille de concordance » entre les valeurs propres et les valeurs-indices. Dans ce but il fixe d'abord, pour chaque critère, un seuil de périphérisation, c'est-à-dire un seuil au-dessus duquel il semble que l'on entre vraiment dans l'agglomération. Pour 13 critères, la valeur-seuil est celle de l'ancien arrondissement de Bruxelles, moins la capitale ; pour d'autres, la valeur-seuil est la moyenne nationale ; pour d'autres encore, cette valeur est choisie d'une façon plus subjective.

En somme on se donne, pour chaque critère, une valeur moyenne de référence ; à cette valeur moyenne on attribue la note-indice 100. C'est-à-dire que si une commune obtient sur le critère 1 des valeurs absolues semblables à celle de l'ancien arrondissement de Bruxelles moins la ville, on lui donne la note-indice 100. L'échelle est maintenant bien accrochée au centre, mais reste à savoir quelle longueur on va lui donner, c'est-à-dire au niveau de quelle valeur absolue on va placer l'indice 0.

Pour la plupart des critères, R. Mols donne l'indice 0 à la valeur absolue 0 ; par exemple une balance migratoire nulle pour les années 1954-1959 se voit appliquer l'indice 0. Dans d'autres cas où les valeurs absolues les plus faibles ne sont pas égales à zéro, on fixe la place du zéro approximativement en fonction de « la nature de chaque critère et de diverses considérations statistiques telles que l'écart des valeurs observées, la nécessité d'avoir des échelles assez sensibles, celle de ne pas désavantager une échelle par rapport à une autre et celle de ne pas obliger à des calculs inutilement onéreux ». Malgré de multiples précautions l'auteur note que de forts écarts subsistent entre les différentes échelles de critères. Dans certains cas, les valeurs les plus élevées dépassent le double ou le triple de la valeur-seuil tandis qu'ailleurs elles ne la dépassent qu'exceptionnellement de 50 %. Ce manque de correspondance entre les échelles revient en fait à donner plus de poids à certains et fausse l'indice global.

Quoi qu'il en soit, toutes les valeurs absolues sont maintenant transformées en valeurs-indices et s'échelonnent entre 0 et 200 ou 300 en passant par la valeur-seuil 100. Il devrait suffire, en principe, d'additionner, pour chaque commune, les indices obtenus sur chaque critère pour obtenir l'indice global de périphérisation de chaque commune. En réalité R. Mols estime nécessaire encore d'avantager certains indices synthétiques par rapport à d'autres car plus représentatifs du phénomène de périphérisation ; il augmente alors l'indice de 10 de ces critères synthétiques de 50 %. Il opère donc une deuxième pondération. Il peut enfin parvenir à un indice général. Chaque commune est classée selon la valeur de son indice dans une des six classes suivantes : localités périphériques de 1^{re} classe (indice moyen : 150 et plus) ; localités périphériques de 2^e classe (indice moyen : 125 à 149) ; localités périphériques de 3^e classe (indice moyen : 100 à 124).

Ces trois types de communes sont considérés comme faisant partie de l'agglomération. Les types suivants n'en font plus partie, mais sont plus ou moins urbanisés : localités marginales supérieures (indice moyen : 85 à 99) localités marginales inférieures (indice moyen : 70 à 84) ; localités submarginales (indice moyen : jusqu'à 69).

Cette typologie se traduit très facilement par une carte synthétique.

Quelles conclusions tirer, en particulier de l'étude de R. Mols ? Les techniques classiques ont l'avantage de ne mettre en œuvre que des méthodes statistiques extrêmement simples et de ne nécessiter qu'une certaine patience et un matériel restreint. Elles présentent cependant trois inconvénients majeurs. Dès le moment où l'on veut jouer sur un nombre de critères importants et couvrir une vaste surface, le travail devient colossal. La longue étude de R. Mols porte sur 90 communes seulement ; comment mener à bien une telle tâche, avec de tels moyens, pour un espace national ou même régional où entrent en jeu plusieurs milliers de communes ? Le deuxième inconvénient provient de ce que le chercheur se voit à plusieurs reprises obligé à des choix subjectifs : première pondération sur les critères de départ, détermination de la valeur-seuil et de la valeur 0, deuxième pondération entre les critères synthétiques. Plusieurs chercheurs travaillant chacun isolément sur le même terrain feraient-ils les mêmes choix ? L'auteur lui-même en est bien conscient qui insiste à plusieurs reprises sur la subjectivité de ses choix, sur l'impossibilité d'éliminer toute part arbitraire ». Enfin, dernier inconvénient, est-on bien sûr d'avoir le droit d'additionner ainsi des indices ? C'est ce que conteste J.-B. Racine qui reproche à certains de « noyer dans un être mathématique unique des phénomènes dont les distributions spatiales respectives ne se recouvrent pas ou s'excluent, assimilant abusivement des phénomènes indépendants »¹².

¹² J.-B. RACINE, « Modèles graphiques et mathématiques en géographie humaine », La Revue de Géographie de Montréal, Les Presses de l'Université de Montréal, 1971, XXV (4), pp. 328-358 ; 1972, XXVI (1), pp. 7-34.

Certaines méthodes mathématiques et l'ordinateur permettent, aujourd'hui, de surmonter une bonne partie de ces difficultés.

Analyse Factorielle et Typologie

L'ordinateur et l'analyse factorielle semblent en effet pouvoir résoudre deux problèmes essentiels : manœuvrer une masse de documentation considérable, condenser et synthétiser cette documentation en perdant le moins possible d'information et en limitant au maximum l'intervention subjective.

L'analyse factorielle, utilisée d'abord par les psychologues, largement employée par les géographes américains ou scandinaves, commence à être bien connue des géographes français. Notre propos sera ici non pas d'expliquer ces techniques — notre bagage mathématique ne nous le permettrait d'ailleurs pas — mais de tenter de voir en quoi elles résolvent certains problèmes qui se révélaient insolubles par les techniques classiques. Nous nous inspirerons largement, pour la première partie de ce qui suit, de l'exposé de J.-B. Racine sur les « Modèles graphiques et mathématiques en sciences humaines »¹³.

Les techniques de base

On part donc, comme plus haut, d'un certain nombre de critères dont on estime qu'ils peuvent servir de base à une typologie de l'espace étudié. Soit une étude sur l'urbanisation des communes rurales d'une région ; on va sélectionner les 20, 30, ou 40 critères qui peuvent varier en fonction de la plus ou moins grande urbanisation de la région, puis, pour chaque commune, rassembler les 20, 30, ou 40 valeurs prises par ces critères. On peut ainsi ordonner une matrice d'information spatiale, c'est-à-dire qu'on associe à chaque commune (que l'on appellera unité d'observation ou individu ou sujet) qui est portée sur l'axe des y, les valeurs prises par différents critères (on dira variable dépendante ou attribut ou caractère) qui sont portés sur l'axe des x.

Jusque-là rien de très nouveau après tout. Nous avons vu qu'à ce niveau R. Mois opérait un premier regroupement de critères puis tentait de rendre comparables ses données. L'analyse mathématique passe tout de suite à cette dernière étape, c'est-à-dire à la standardisation de la matrice. La valeur de chaque attribut est remplacée « par le rapport existant entre les deux écarts fondamentaux qui caractérisent sa position dans la distribution et l'écart-type de la distribution ». Toutes les moyennes sont ainsi égales à 0 (c'est l'équivalent de l'indice 100 de R. Mois) et dans la plupart des cas les cc notes » varient entre la moyenne et ± 3 écarts-types. On a donc obtenu, avec des moyens simples et relativement rapides, des valeurs qui se réfèrent à une même échelle et qui sont beaucoup plus homogènes que celles auxquelles on parvenait pour la périphérie bruxelloise.

¹³ Toutes les citations de ce sous-chapitre sont de J.-B RACINE, op. cit.

Ce stade atteint, l'analyse classique se contente — après pondération parfois — d'additionner, pour chaque unité d'observation, les valeurs prises par les différents critères en ce lieu. Or certains des critères sont peut-être, en partie, le reflet de réalités identiques ; peut-être même deux critères ont-ils des distributions semblables, ou à peu près, et donc font-ils double emploi ? L'analyse classique n'a aucun moyen rapide d'éliminer ces doubles comptes, sauf à superposer deux à deux les cartes obtenues pour chacun des critères pour voir si certaines ne sont pas semblables ; mais dès que les variables sont nombreuses c'est un gros travail qui, quoi qu'il en soit, ne débouche pas sur des précisions chiffrées. L'analyse factorielle va, entre autres choses, régler ce problème. Deux méthodes essentielles sont utilisées : l'analyse en composantes principales, l'analyse des correspondances.

L'analyse en composantes principales

Si l'on considère deux variables, on appelle coefficient de corrélation linéaire la mesure du degré de liaison entre leurs variations ou, si l'on veut, « du degré de ressemblance entre chacune des deux distributions ». Ce coefficient oscille entre + 1, lorsque les valeurs des distributions sont identiques et - 1, lorsque celles-ci sont opposées, en passant par 0 si les variables sont indépendantes. On dresse ainsi, en utilisant uniquement les variables, une matrice de corrélation qui donne le tableau des corrélations entre toutes les variables prises deux à deux. Ce tableau permet de préciser le système des liaisons entre les variables et donc de juger des redondances possibles de l'une à l'autre. En schématisant, disons que l'on a une mesure du degré de ressemblance des variables entre elles. Si l'on arrêta à l'élaboration mathématique on obtiendrait déjà un résultat fort supérieur à celui de l'analyse classique : il serait dès à présent possible de mettre sur pied un système objectif de pondération des critères puisqu'on sait ce que chacun a en commun avec tous les autres.

L'analyse factorielle va permettre de faire plus, c'est-à-dire de réduire le support de l'information en mutilant celle-ci le moins possible. On suppose en effet que, derrière un certain nombre de variables corrélées, se cache un facteur commun qui décrit, justement, une bonne part de cette corrélation entre les variables. L'analyse factorielle permet cette détermination de facteurs communs. Elle sort, par ordre décroissant, les composantes, c'est-à-dire les différents facteurs ; on ne retient de ceux-ci que les plus importants, c'est-à-dire ceux qui fournissent, en quelque sorte, un pourcentage maximum « d'explication ». Elle donne également les saturations, c'est-à-dire les coefficients de corrélation entre chacune des variables et chacun des facteurs ou, en schématisant encore, le pourcentage « d'explication » de chaque variable par chacun des facteurs. A ces facteurs, il faut donner -une signification, un nom, en fonction des variables essentielles qui y entrent en combinaison.

On voit donc comment, sans intervention subjective, on est passé de variables non comparables à des variables ramenées à une même échelle et comment, à partir

de multiples variables, on est arrivé à quelques facteurs synthétiques dont — au surplus — on connaît bien la composition. On sait également l'importance de l'information perdue en cours de route.

Cette connaissance des facteurs latents suffit à beaucoup d'utilisateurs, en particulier au psychologue. Suffit-elle au géographe ? Dans un premier temps, oui. Il peut faire une cartographie, déjà à demi synthétique, des quelques facteurs qu'il a découverts. En effet l'ordinateur lui donne la note obtenue, sur chaque unité d'observation, par chacun des facteurs. Il peut faire ainsi autant de cartes qu'il a de facteurs. Ces cartes sont peu nombreuses — moins nombreuses, quoi qu'il en soit, que celles que l'on peut faire avec toutes les variables de départ — et elles peuvent déjà permettre une comparaison visuelle extrêmement intéressante.

Cependant le géographe désire, dans un deuxième temps, dépasser les synthèses partielles pour aboutir à une synthèse globale qui tienne compte, pour chaque unité d'observation, du poids respectif de chacun des facteurs. Additionner, pour chaque observation, les scores obtenus sur les différents facteurs, reviendrait à tomber sous le coup des critiques faites aux méthodes classiques, c'est-à-dire à « assimiler des phénomènes indépendants ». L'analyse mathématique va donc essayer de classer les unités d'observation en fonction des variations des facteurs qui les décrivent, de regrouper, pour en faire des types, des unités dont les facteurs varient à peu près de la même façon.

L'idée générale est celle que nous retrouverons dans l'analyse des correspondances. Chaque individu est représenté, dans un espace multidimensionnel, par un vecteur — ou le point-extrémité de ce vecteur — dont on peut assimiler les coordonnées aux valeurs des différents facteurs reconnus auparavant par l'analyse factorielle. Dans cet espace, chaque individu a une certaine place. Il devient donc possible « de mesurer les liens de proximité entre chacune des observations et de les comparer, de façon à définir quelles sont les observations les plus proches les unes des autres ». On va donc tenter de classer ces unités d'observation en un certain nombre de groupes, en fonction de leur plus ou moins grande proximité dans l'espace multidimensionnel et aboutir à une typologie.

Il semble bien, cependant, que l'outil mathématique soit, ici, beaucoup moins efficace. De nombreuses méthodes sont utilisées ; aucune ne donne, à elle seule, pleinement satisfaction. On est obligé, le plus souvent, d'utiliser plusieurs techniques de classement qui permettent, par approches successives, d'homogénéiser au maximum les groupes. Le chercheur est obligé de faire des choix à caractère plus ou moins subjectif car le classement automatique aboutirait parfois à mettre dans certaines classes des unités qu'une bonne connaissance du terrain devrait faire placer ailleurs. Et J.-B. Racine rappelle le mot de Johnson sur « le caractère décidément subjectif des méthodes objectives quand il s'agit de choisir entre tous les choix possibles en matière de classification et de recherche typologique ».

Même s'il reste des incertitudes méthodologiques à la dernière étape, on peut affirmer, cependant, que ces techniques ont abouti à un certain nombre de résultats prometteurs en matière de typologie. Nous n'avons pas encore connaissance des résultats des travaux de certains Ateliers régionaux d'aménagement rural qui ont tenté de faire des typologies de l'espace rural. En revanche, de nombreux travaux ont déjà été effectués sur l'espace urbain, en particulier aux États-Unis¹⁴. En France même, dès 1965, les Cahiers de l'institut d'aménagement et d'urbanisme de la Région Parisienne publiaient une « Comparaison et classification des communes de l'agglomération parisienne »¹⁵. A partir de 69 critères appliqués à 251 communes de l'agglomération parisienne, l'IAURP a reconnu 45 facteurs indépendants. De ces 45 facteurs on en a retenu 6 que l'on a appelés : équilibre de l'emploi, urbanisation et industrialisation, migrations sur Paris, niveau de vie, structure démographique, évolution démographique, qui sont chacun cartographiés. En fonction de la combinaison de ces facteurs, un programme de classement séquentiel a distingué 13 groupes de communes et a permis une carte de synthèse. Depuis, M. Roncayolo¹⁶ et E. Dalmasso¹⁷ ont utilisé ces techniques pour des travaux sur Marseille et Milan.

Un deuxième type d'analyse factorielle est également employé. C'est l'analyse des correspondances. On va voir qu'elle rend des services et qu'elle pose des problèmes un peu différents.

L'analyse factorielle des correspondances

On part de l'idée, exposée plus haut à propos des facteurs, que chaque unité d'observation (et non plus chaque facteur) est représentée dans un espace multidimensionnel par un point dont on peut assimiler les coordonnées aux valeurs des différents caractères. Chaque unité a donc une certaine place dans cet espace et l'ensemble forme un nuage de points. Ce nuage de points, dans un espace multidimensionnel, n'est évidemment pas représentable pour nous qui vivons en trois dimensions. Le but de l'analyse des correspondances va donc être double.

D'une part, il faut ramener ce nuage de points dans un espace à plusieurs dimensions à un nuage de points dans un espace à un nombre réduit de dimensions, tout en conservant, autant que faire se peut, la forme du nuage, donc en perdant le moins possible d'information. D'autre part, on ne veut pas seulement simplifier l'espace de représentation, mais également tenter d'expliquer l'organisation du nuage. Cette explication, on la cherche en analysant les principales directions d'allongement du nuage. En effet, la répartition des variables le long de ces axes

¹⁴ Cf. par exemple Brian J. L. BERRY & Frank E. HORTON, *Geographic Perspectives on Urban Systems*, Prentice Hall, Englewood Cliffs, NJ, 1970.

¹⁵ INSTITUT D'AMÉNAGEMENT ET D'URBANISME DE LA RÉGION PARISIENNE, « Comparaison et classification des communes de l'agglomération parisienne », *Les Cahiers de l'IAURP*, 1965, 3, pp. 1-102.

¹⁶ M. RONCAYOLO, communication du 4 décembre 1971 à l'Association des géographes français.

¹⁷ E. DALMASSO, « Essai d'application de quelques méthodes statistiques à la région milanaise », *Bulletin de l'Association des géographes français*, 1971, 393-394, nov.-déc., pp. 377-392.

permet généralement de donner un sens à ces allongements, de retrouver, en quelque sorte, les facteurs qui organisent le nuage dans tel ou tel sens.

On voit donc que l'on parvient, comme dans l'autre type d'analyse factorielle, à réduire le support de l'information et à dégager des facteurs dont on peut mesurer les poids respectifs dans « l'explication » globale. Mais cette méthode a l'avantage de permettre de travailler à la fois sur les individus et sur les caractères. En effet on peut placer, dans le même espace, caractères et individus ; on peut donc savoir quels caractères sont les plus proches de tel ou tel individu, donc savoir « qui a quel caractère »¹⁸. Dans la pratique, le passage à l'ordinateur fournit, pour chaque caractère et chaque individu, les coordonnées par rapport aux axes essentiels du nuage tels que les a définis l'analyse. On dresse, à partir des coordonnées des caractères, des graphiques où sont représentés, dans un espace à deux dimensions, les axes 1 et 2, puis l'axe 3 par rapport au plan des axes 1 et 2 et éventuellement d'autres axes, dans la mesure où ceux-ci peuvent être significatifs. On dresse ensuite, pour les mêmes axes, d'autres graphiques mais où apparaissent, cette fois, les individus. Enfin on passe à des graphiques synthétiques où, sur les mêmes axes, on positionne à la fois les individus et les caractères. On peut alors répondre au « qui a quel caractère ». L'analyse de l'organisation du nuage permet de donner un nom aux axes, c'est-à-dire aux facteurs.

Reste la dernière étape, celle du classement des individus en catégories selon la combinaison, sur chacun d'eux, des différents facteurs. Ici nous retrouverons des difficultés analogues à celles que nous avons rencontrées avec l'autre type d'analyse et qui semblent inhérentes à toute typologie.

On utilise les graphiques qui représentent les individus sur les axes 1 et 2 d'abord. Si le nuage de points éclate, en fait, en plusieurs masses séparées par des blancs au moins relatifs, pas trop de problèmes. On divisera les individus en autant de classes qu'il y a de masses de points. L'analyse de l'axe 3 permettra éventuellement de recouper ces nuages selon une troisième dimension, donc de multiplier le nombre de classes et de raffiner la typologie. Les graphiques mêlant caractères et individus permettront de définir les spécificités de chaque classe. Cependant il sera bien difficile, sinon impossible, d'utiliser les autres axes pour cette typologie.

Si le nuage est massif, ou du moins s'il ne comporte pas de masses bien typées, on se trouve contraint à trancher dans le vif et à définir, un peu a priori, par rapport aux différents axes, des zones de valeurs forte, moyenne, faible qui permettent un découpage en classes. Ces catégories sont utilisables en ce sens qu'elles regroupent bien les individus selon leur place dans l'espace défini, mais il est évident que les classes risquent fort d'intégrer des individus un peu marginaux par rapport à

¹⁸ J.-P. MASSONIE, D. MATHIEU et J.-C. WIEBER, « Application de l'analyse factorielle à l'étude des paysages », Cahiers de Géographie de Besançon (Séminaires et notes de recherche), 1971,4, sept., pp. 1-51.

l'ensemble. C'est, nous semble-t-il, une des faiblesses de cette méthode. Mais, par ailleurs, son caractère puissamment synthétique en fait un instrument d'analyse de premier ordre. Cette méthode a été utilisée ici même par F. Loux et M. de Virville¹⁹ à propos de l'étude du système social du Châtillonnais ; c'est celle qui a été pratiquée plusieurs fois déjà, mais pour d'autres types de recherche, sous la direction de J.-P. Massonie, à l'Institut de géographie de Besançon²⁰ ; c'est celle que nous pensons utiliser pour une étude typologique des communes rurales de Franche-Comté.

Éléments d'un bilan

Rappelons rapidement quelques-uns des avantages de ces méthodes mathématiques pour une analyse typologique : prise en compte d'une vaste information et d'un grand nombre d'unités d'observations ; comparabilité de cette information ; réduction systématique, avec une perte minimum et chiffrable, du support de l'information ; connaissance chiffrée du niveau « d'explication » de chacun des facteurs ; cartographie simple et systématique des variables, naturellement, mais aussi des facteurs et de la combinaison de ces facteurs ; dans une certaine mesure, typologie synthétique et raisonnée de l'espace.

Cependant ces méthodes ont leurs limites. Nous ne parlerons pas de celles qui sont inhérentes à la quantification elle-même : les méthodes classiques quantifient elles aussi. L'analyse des corrélations, l'analyse factorielle ne posent guère de problèmes en elles-mêmes ; encore faut-il bien voir qu'une forte corrélation entre deux variables ne signifie pas forcément une relation causale entre elles. Les limites se situent surtout, on l'a vu, au niveau de la typologie synthétique. Il est vrai qu'aucune classification n'est jamais totalement satisfaisante et qu'il y a toujours des individus difficilement classables qui obligent à trancher dans le vif.

Quoi qu'il en soit, l'utilisation de telles méthodes n'élimine pas, bien au contraire, l'expérience géographique classique, en particulier celle du terrain. Rappelons que cette expérience intervient à plusieurs stades : choix des variables, caractérisation des facteurs, fixation de la typologie, interprétation des cartes. Nous ajouterons que ces méthodes nécessitent également une compréhension sans cesse plus profonde de la nature même de la géographie.

Les limites de la notion de type

Nous terminerons sur une interrogation et sur une mise en garde. La géographie, comme science de l'organisation de l'espace humanisé, trouve un de ses thèmes d'étude majeur dans des essais de classification, de typologie de l'espace. Il faut se souvenir cependant que toute typologie n'est que partielle puisqu'elle ne peut

¹⁹ F. LOUX et M. de VIRVILLE, « Le système social d'une région rurale, le Châtillonnais », *Études rurales*, 1969, 35, juil.-sept., pp. 5-135.

²⁰ En particulier, J.P. MASSONIE, D. MATHIEU et J.-C. WIEBER, op. cit.

prendre en compte, même avec des moyens modernes, toutes les variables qui sont impliquées dans les phénomènes que la géographie cherche à expliquer. Elle fait donc un choix, plus ou moins entaché d'a priori au départ : toute typologie est donc partielle et partielle.

S'il en est ainsi, est-on sûr que les variables utilisées permettent de définir, de limiter des êtres géographiques vraiment vivants, des configurations spatiales vécues économiquement, socialement, psychiquement par les hommes ? Il faut bien voir que, dans une typologie, moins on prend de variables plus on détermine des êtres géographiques frustes, certes, mais nets ; plus on inclut de variables plus on définit des configurations riches de sens mais abstraites.

Cela ne signifie pas qu'il faille renoncer aux typologies mais il faut savoir qu'on ne peut y trouver que ce que l'on y met, c'est-à-dire une réalité partielle. Ce ne sont que des instruments de recherche dont il faut vérifier la validité, sur le terrain en particulier.