

Écologie factorielle et attributs géographiques

Jean-Bernard Racine and Marc Cavalier

Volume 16, Number 38, 1972

URI: <https://id.erudit.org/iderudit/021054ar>

DOI: <https://doi.org/10.7202/021054ar>

[See table of contents](#)

Publisher(s)

Département de géographie de l'Université Laval

ISSN

0007-9766 (print)

1708-8968 (digital)

[Explore this journal](#)

Cite this article

Racine, J.-B. & Cavalier, M. (1972). Écologie factorielle et attributs géographiques. *Cahiers de géographie du Québec*, 16(38), 213–241. <https://doi.org/10.7202/021054ar>

Article abstract

The authors analyse, one after the other, the algebraic and theoretical constraints of currently applied algorithms of factorial ecology as used by geographers. The continuous ambiguity of most results led them to point out that mathematical procedures should always be submitted to geographical interpretation. However the mathematical apparatus implements the critical checking and improves geographical results. The authors then proceed to the problem of their « optimality », with special reference to the method of discriminant analysis. Finally, they stress the concept of geographical indicators progressive discovered through factorial and discriminant iterations ; these indicators may seem and be simple, but they are loaded with a real factorial pedigree.

ÉCOLOGIE FACTORIELLE ET ATTRIBUTS GÉOGRAPHIQUES

par

Jean-Bernard RACINE et Marc CAVALIER

Département de Géographie, Université d'Ottawa

Cet essai s'inscrit dans le cadre d'une recherche à long terme sur la mise au point d'un modèle de recherche spécifique à une géographie qui se présente comme une science dynamique de la différenciation et de l'organisation des éco-systèmes spatiaux ¹.

Nous pensons qu'il n'est pas inutile, malgré le récent ouvrage édité par Brian J.L. Berry ² de donner ici quelques exemples des limites, des dangers, mais aussi de l'extraordinaire fécondité de l'utilisation de l'analyse factorielle en géographie, en prenant comme thème de réflexion celui de la recherche des indicateurs socio-économiques les plus adéquats, géographiquement et mathématiquement parlant, pour une utilisation ultérieure dans l'analyse typologique et dans la modélisation des éco-systèmes spatiaux.

I. LA RÉDUCTION D'UNE MATRICE D'INFORMATION ET LA CRÉATION MATHÉMATIQUE D'ATTRIBUTS SPATIAUX : TROIS PROBLÈMES FONDAMENTAUX

Une fois admis le principe de la nécessité d'une première réduction factorielle de l'ensemble des caractéristiques recueillies pour définir chacune des unités d'observation formant les différents éléments constitutifs du système spatial analysé ³, et compte tenu des contraintes mathématiques liées à ce type d'analyse (ainsi par exemple le fait que le nombre des variables doit obligatoirement être égal ou inférieur au nombre des unités d'observation), le problème fondamental qui se pose au géographe réside dans le degré de confiance qu'il peut mettre dans les êtres mathématiques nouvellement créés par cette réduction.

¹ RACINE, J. B. (1972), *Écologie factorielle et écosystème spatiaux*. Dans George E. BOURGOIGNIE (ed.) *Perspectives en écologie humaine*. Paris, Presses Universitaires de France, (sous presse).

² BERRY, B. J. L., Guest Ed. (1971) *Comparative Factorial Ecology. Economic Geography*, 47 (2), (1971 suppl.), 367 p.

³ RACINE, J. B. (1971), *Modèles graphiques et modèles mathématiques en géographie humaine*, I. *Rev. Géogr. Montr.*, XXV (4) : 323-358.

1. Une dépendance abusive : poids locaux et attributs spatiaux

On sait qu'à cette fin chacune des « p » colonnes de la matrice des saturations est habituellement multipliée par chacune des valeurs standardisées des attributs de l'observation « i », de façon à obtenir un nombre p de poids locaux pour cette observation. Lorsque le modèle factoriel de décomposition de la matrice de corrélation n'est pas celui du « PCA » (Principal component analysis) mais plutôt celui du « PAFA » (Principal axis factor analysis) par exemple, ou encore de « l'analyse-image » (« image analysis »), la solution n'est pas si simple. Harman⁴ et d'autres ont pourtant trouvé des solutions qui sont aujourd'hui largement utilisées, telle celle mise au point par Crawford⁵ dans sa thèse de psychologie consacrée à l'analyse-image qu'après Greer-Wootten⁶ nous utilisons systématiquement⁷. Mais quelle que soit la solution utilisée, le principe de fond est toujours le même et les poids locaux dépendent toujours en dernière analyse de l'ensemble des attributs spatiaux originaux et du jeu de chacune des saturations obtenues dans la matrice factorielle, alors que le facteur (ou la composante) n'est identifié et habituellement défini qu'à partir des seules variables ayant obtenu sur le facteur considéré des saturations significativement élevées, tant sur le pôle négatif que sur le pôle positif, le seuil de 0,40 étant le plus souvent retenu.

Il y a cependant entre les poids locaux des facteurs et les attributs spatiaux originaux une dépendance extrêmement dangereuse, comme Racine en fait la démonstration dans son troisième article sur les « modèles graphiques et mathématiques en géographie humaine »⁸, dépendance qui peut même interdire à toutes fins pratiques l'utilisation des poids locaux d'un facteur ou d'une composante principale pour décrire adéquatement les variations spatiales d'une dimension latente qui n'est définie au plan conceptuel qu'à partir des plus fortes saturations, positives ou négatives. Il suffit en effet d'une seule valeur exceptionnelle dans une distribution pour que même pondérée par une saturation très faible, une quelconque observation vienne se ranger dans le continuum des poids locaux, à une place qui est contredite par les attributs ayant servi à définir la dimension que l'on veut représenter spatialement. Le même phénomène peut au contraire contribuer à « noyer », par le seul jeu des signes positifs ou négatifs, une observation dans la partie médiane du continuum des scores alors qu'elle possède de fortes valeurs sur les attributs retenus dans la définition du facteur, voire même les plus fortes valeurs.

4 HARMAN, H. H. (1966), *Modern Factor Analysis*. Chicago, University of Chicago Press. 2ème édition, 474 p.

5 CRAWFORD, C. B. (1966), *Analytic Methods of Rotation in the Determination of the Number of Factors*. Montréal, McGill Univ., Ph.D Thesis.

6 GREER-WOOTTEN, Bryn (1972), Changing Social Areas and the Intra-Urban Migration Process. *Rev. Géogr. Montr.*, XXVI (3) : 271-292.

7 RACINE, J. B. (1972), Modèles graphiques et modèles mathématiques en géographie humaine, II. *Rev. Géogr. Montr.*, XXVI (1) : 7-34.

8 RACINE, J. B. (1972), Modèles graphiques et modèles mathématiques en géographie humaine, III. *Rev. Géogr. Montr.*, XXVI (3) : 321-332.

Une partie de la solution à ce problème clé de l'interprétation des facteurs et de la validité des résultats chiffrés pour les analyses subséquentes (analyses typologiques, utilisation des « scores » dans des modèles de régression par exemple) peut être cherchée dans l'utilisation systématique du procédé VARIMAX de Kayser⁹, voire de la normalisation des distributions au départ de l'analyse, normalisation qui est de toutes façons requise pour tous les types d'analyse factorielle autres que celui proposé par Hotelling¹⁰. Mais ni l'une ni l'autre de ces deux manipulations n'efface complètement le problème, même si elles contribuent fortement à amoindrir les effets des redondances ou autres « bruits de fonds » mathématiques. À l'expérience, il faut bien constater les dangers d'une utilisation directe des résultats d'une analyse factorielle qui ne référerait pas aux valeurs originales, qui n'impliquerait pas une confrontation continue de ces valeurs — et mieux encore, des valeurs standardisées — aux valeurs des saturations d'une part, à l'échelle des poids locaux d'autre part. Seul en effet un retour aux valeurs des variables d'entrée permet de distinguer quel est ou quels sont le ou les attributs spécifiques (ayant obtenu de fortes saturations positives ou négatives) responsables de la position de chacune des observations sur le continuum des poids locaux, voire même de distinguer lequel a été le plus important dans le calcul de la valeur finale.

Cette dépendance entre les valeurs de *l'espace factoriel* et celles de *l'espace des attributs* augmente évidemment avec le nombre de variables retenues à l'origine pour constituer la matrice d'information spatiale qui sera soumise à l'algorithme factoriel, puisque ce sont justement les saturations, même très faibles, qui font problème lorsque l'on cherche à cartographier ou à utiliser dans une analyse subséquente les êtres mathématiques « conceptuels » nouvellement créés. De là l'intérêt d'une réflexion théorique préalable sur la nature des variables à introduire au départ dans l'analyse, réflexion géographique dans la mesure où elle essaiera d'évaluer une double série de rapports, les rapports que chacune des variables entretient avec le problème analysé, les rapports que chacune des variables entretient avec toutes les autres. La constitution préalable d'une matrice des corrélations simples existant entre toutes les variables dont on imagine au départ, pour une raison ou une autre, l'emploi dans une analyse factorielle, vient évidemment féconder une telle réflexion théorique, en lui apportant une grande efficacité opérationnelle. Or une *contrainte* du modèle de calcul matriciel utilisé pour l'analyse factorielle — le fait qu'on ne puisse associer aux unités d'observation retenues dans la matrice plus de variables qu'il n'y a d'observations ($n \geq m$) — oblige le géographe à un double effort, de réflexion théo-

⁹ KAYSER, H. F. (1958), The Varimax Criterion for Analytic Rotation of Factor Analysis. *Psychometrika*, 23 : 187-200.

— (1959), Computer Program for Varimax Rotation of Factor Analysis. *Educ. and Psychol. Measurement*, 19 : 413-420.

¹⁰ HOTELLING, H. (1933), Analysis of a Complex of Statistical Variables into Principal Components. *Journ. Educ. Psychology*, 24 : 417-441, 498-520.

rique d'une part, et de filtrage progressif de l'information d'autre part. Ceci conduit à un découpage préalable de l'information originale en un certain nombre de trames * logiques dont on sait que la superposition à la trame des organisations naturelles définit justement cette catégorie de faits que l'on appelle les faits géographiques. À l'expérience cependant, cette « contrainte » se révèle extrêmement féconde et vient résoudre en grande partie le problème posé par la plus ou moins grande pertinence des poids locaux des facteurs.

2. *Les différentes trames de l'organisation spatiale*

On conviendra que la première étape de toute analyse géographique réside bien dans la description, la définition, le classement et la typologie des différentes formes d'organisation de la vie humaine à la surface de la terre, — ou dans le cadre spatial étudié « E » — telles que traduites par les *paysages géographiques*. Or ces paysages résultent de la superposition des différentes trames des organisations naturelles de l'espace physique et des trames humaines, visibles ou invisibles, à zéro, une, deux ou n dimensions, d'ordre statique ou d'ordre dynamique, ponctuelles ou relationnelles, selon lesquelles se découpe l'espace économique et social, cet espace humain qu'il convient d'*explorer* dirait William Bunge¹¹. On doit à Ph. Pinchemel¹² un premier essai d'identification de ces trames (trame d'appropriation, trame d'organisation, trame d'utilisation et de fonctionnement). Mais les travaux des spécialistes de l'écologie factorielle ont montré qu'on pouvait — et qu'il fallait — concevoir d'autres découpages et que par exemple l'espace social des grandes villes américaines se structurait d'une façon très générale en fonction de la superposition de trois trames indépendantes : la trame socio-économique, la trame socio-démographique et la trame socio-ethnique.

Le modèle n'est pourtant pas universel. L'étude de Pedersen¹³ sur Copenhague a révélé que si les deux premières dimensions de l'écologie factorielle de cette ville étaient comme en Amérique du Nord, le « statut socio-économique » et le « statut familial », la dimension « croissance et mobilité de la population » remplaçait le facteur ethnique de « ségrégation », comme découvert en Amérique. C'est le même résultat qu'on obtient d'ailleurs si on change l'échelle de l'analyse en Amérique du Nord et si le test porte par exemple sur la seule couronne urbaine. Appliqué à Helsinki par Sweetser¹⁴, le même test a prouvé en outre que dans cette ville le facteur ethnique

* Peut-être conviendrait-il de parler de *familles d'attributs spatiaux* plutôt que de trames.

¹¹ BUNGE, W. (1966), *Theoretical Geography*. Lund, Gleerup, 289 p. « Lund Studies in Geography », ser. C, no. 1.

¹² PINCHEMEL, Ph. (1968), Redécouvrir la Géographie. *Annales de l'Université de Paris*, (3) : 350-360.

¹³ PEDERSEN, P.O. (1967), An Empirical Model for Urban Population Structure. *Proceedings of the First Scandinavian — Polish Regional Science Seminar*. Varsovie, Polish Scientific Publisher.

¹⁴ SWEETSER, F. (1965), Factorial Ecology : Helsinki 1960. *Demography*, 1 : 372-386.

n'était pas indépendant du facteur socio-économique, la minorité suédoise étant nettement plus aisée que la moyenne et étant de ce fait plus à même de contrôler la localisation de son habitat au sein de la cité. Surtout, les études conduites en pays encore en voie de développement ou soumis à un processus colonial, comme l'étude de Janet Abu-Lughod¹⁵ sur le Caire ou de Berry et Rees sur Calcutta¹⁶ ont montré que les dimensions de base de l'écologie factorielle pouvaient n'avoir dans ces villes situées dans un tout autre contexte historique, culturel économique et social, que de très lointains rapports avec celles des villes nord-américaines.

Quoi qu'il en soit, à partir de l'inventaire aussi complet que possible, sinon exhaustif, de tous les éléments intervenant dans les rapports que le géographe doit étudier, on peut concevoir une mise en ordre de l'information, trame par trame, chacune de ces dernières pouvant constituer des matrices d'information spatiales d'ordre $n \times m$, dans lesquelles m est toujours égal ou inférieur à n , qui tiennent compte de la nature des faits géographiques, et de la nature propre de leurs attributs spatiaux : attributs ressortissant au domaine naturel, aux données de l'utilisation du sol, du logement et des conditions de la démogéographie (démographie, statut familial, statut socio-économique), des activités économiques et de gestion (activités industrielles, activités commerciales, activités de service, finances municipales). C'est déjà là une première classification qui, elle non plus, n'a évidemment rien d'universel, mais qui peut être prise comme un point de départ provisoire permettant de réduire le nombre des attributs sur lesquels on fera porter les premières analyses, et de chercher à mettre en évidence leurs relations structurelles spécifiques, pour mieux comprendre leurs modalités d'agencement, dans un contexte phénoménologique homogène, l'utilisation du sol ou les pyramides d'âge par exemple.

On évitera ainsi, dans une grande mesure, les pièges des « bruits de fonds » mathématiques qui masquent l'essentiel de ce que l'on cherchait à mettre en évidence par l'analyse factorielle ; on sera beaucoup plus à l'aise pour contrôler étape par étape la transformation de l'information et ses résultats par la réflexion théorique ; on pourra surtout dégager progressivement, de la suite d'analyses auxquelles seront soumises chacune des trames, les indicateurs les plus synthétiques et les plus représentatifs de l'ensemble de l'information retenue au départ, pour finalement les combiner dans une matrice d'information réduite, au sein de laquelle chacun des attributs aura un véritable « pedigree » et l'analyse finale un arbre généalogique précis. À la limite, et c'est l'ambition de cet essai, les attributs retenus pour la dernière analyse peuvent être de simples variables, non composites, représentatives des facteurs de la différenciation spatiale, véritables *indicateurs* géographiques directement utilisables par n'importe quel utilisateur.

¹⁵ LUGHOD, J. Abu (1969), Testing the Theory of Social Area Analysis : the Case of Cairo, Egypt. *American Sociological Review*, 34 : 198-212.

¹⁶ BERRY, B.J.L. & REES, H. (1969), The Factorial Ecology of Calcutta. *Am. Journ. of Sociology*, (74) : 445-491.

Les analyses qui suivent montrent comment, à quel prix et avec quel niveau de pertinence scientifique, le recours de manière proprement *itérative* au modèle des composantes principales de Hotelling, permet de passer d'une écologie factorielle à une véritable géographie factorielle, fondée sur la prise en compte d'indicateurs géographiques simples et représentatifs, dont la validité aura pu être testée au niveau des exigences de l'algorithme et du cadre conceptuel utilisé comme au niveau des exigences d'une connaissance proprement géographique.

3. *L'espace factoriel et les exigences conceptuelles et opérationnelles de la notion de système*

Dans l'état des connaissances actuelles il n'est plus nécessaire de définir la notion d'espace factoriel, ni même la notion de système. Elles font maintenant partie de corpus méthodologique et conceptuel de la majorité des géographes, du moins en Amérique du Nord, en Angleterre et en Scandinavie. Il convient cependant de rappeler ici quelques principes de base, ceux-là mêmes qui vont légitimer la méthode de réduction factorielle utilisées ci-après.

a) *Le primat de la totalité*

La notion de *tout* introduit par exemple celle du « primat de la totalité ». Or ce primat de la totalité a ses exigences au plan méthodologique. Si nous pouvons faire des réserves quant à l'application sans nuance de la théorie des systèmes à l'étude des faits géographiques c'est, qu'à notre connaissance du moins, tout en étant un merveilleux instrument de travail théorique capable de guider la recherche, cette théorie doit bien passer par un volet opérationnel : celui, nous l'avons dit, des différentes techniques de l'analyse statistique multivariée. Or toutes ces techniques quantitatives, telles qu'utilisées par les géographes en mal de recherche structurelle ou « systémiste » sont des techniques qui pour l'instant reposent sur une pierre d'angle et une seule : la corrélation linéaire. Le recours à la notion de système en matière d'écologie factorielle repose donc en dernière analyse sur la pertinence, dans chacune des études entreprises, de deux hypothèses fondamentales : que chacun des attributs considérés dans la matrice d'information spatiale originale joue véritablement un rôle dans la définition de la structure du système perçu dans sa totalité, ce que pourra révéler l'analyse dite des communautés ; que chacun des attributs considérés entretienne réellement des relations de colinéarité avec au moins quelques-uns des autres attributs, un au moins, ce que révélera directement l'analyse des corrélations existant entre chacun des attributs pris deux à deux. Dans le cas contraire à quoi bon fonder une recherche sur des méthodes qui impliquent au départ les deux pré-requis mentionnés ? En fait, s'il est inadéquat de fonder une recherche sur ces méthodes, on peut et on doit sans doute toujours les utiliser, dans une première étape, pour tester l'existence de ces relations d'interdépendance liées à la notion même de système. On pourra ainsi écarter des analyses

subséquentes, sur lesquelles se fonderont vraiment les résultats, tous les attributs, (voire au besoin les éléments strictement singuliers qui viennent faire exception à la règle liant entre eux les attributs) qui manifestement n'entrent pas, de façon linéaire du moins, dans le jeu du système et dans la définition de la structure qui lui est liée et qui explique ses comportements possibles.

b) *L'intérêt de l'analyse des communautés*

On sait que les communautés sont calculées le long des rangées d'une matrice factorielle, de la même façon qu'ont été calculées les sommes des saturations le long des colonnes, sommes exprimant le poids ou le rôle joué par chacun des facteurs. La somme des carrés des saturations contenues dans les rangées indique au contraire la communauté de chacun des attributs considérés, c'est-à-dire le poids ou le rôle (qui peut lui aussi se lire en pourcentage) suivant lequel chacun d'eux a contribué à la définition de la matrice factorielle. Plus les communautés sont fortes, plus grand est le rôle (action et réaction) de l'attribut au sein du système analysé. Au contraire, plus cette communauté est faible, plus l'attribut concerné est linéairement indépendant du système de liaisons multiples analysé. C'est alors le signe qu'il faut chercher en dehors du système défini par la matrice d'information spatiale, les causes (ou les concomitances) qui président ou sont liées aux modalités de répartition de cet attribut et les caractéristiques qui peuvent s'y associer. Pour « resserrer » l'analyse, on est donc en droit de décider de laisser cet attribut de côté, puisqu'il ne joue aucun rôle, ne dépendant pas, ou très peu, des autres attributs retenus et n'agissant pas ou très peu sur eux.

Diverses expériences nous ont cependant montré, et l'une d'elle a été publiée par J. B. Racine¹⁷ dans la première partie de son essai sur les « Modèles graphiques et mathématiques en géographie humaine », que la hiérarchie des valeurs des communautés n'est pas forcément conforme à la hiérarchie que manifeste la matrice des corrélations. Ainsi, deux attributs peuvent avoir d'assez faibles communautés tout en entretenant tous deux des corrélations importantes soit entre eux soit encore avec un troisième attribut. L'inverse est vrai également. De même, un attribut peut n'avoir aucune corrélation avec des voisins tout en ayant une forte communauté. Il détermine alors à lui seul l'existence d'une composante principale ou d'un facteur supplémentaire qui disparaît de l'analyse si l'attribut est enlevé de la matrice d'information originale. Inversement encore une fois, un attribut peut posséder une très faible communauté, n'entretenir que de médiocres corrélations avec les autres attributs et obtenir pourtant une très forte saturation sur un des facteurs importants, voire le plus important des facteurs découverts, facteurs également indexés par plusieurs autres attributs. Dans de tels cas, à notre connaissance du moins, il s'agit d'attributs extérieurs au système de relations internes que décrivent les autres attributs, reliant

¹⁷ RACINE, J.B. *Opus cit.*, voir note 3.

par exemple le système étudié à un système plus vaste mais dont le premier fait partie.

c) *Le choix du niveau des colinéarités et les implications de la rotation VARIMAX*

L'analyse des communautés se révèle donc incapable à elle toute seule de légitimer de façon définitive et objective le choix des variables à retenir pour l'analyse factorielle finale puisque les communautés peuvent avoir, en regard du système complet des colinéarités enregistrées, des significations fort différentes. C'est pourquoi ce premier tri effectué au niveau des communautés doit être complété, voire croisé, par les résultats d'un autre tri, celui que l'on effectue directement au niveau de la matrice des corrélations linéaires unissant deux à deux chacun des attributs spatiaux, que ceux-ci soient des variables originales ou déjà des poids locaux composites issus d'analyses factorielles préalables. Rien de plus légitime en effet puisque l'analyse factorielle repose par définition sur le postulat de l'existence de relations linéaires entre différents groupes d'attributs spatiaux, un certain nombre d'entre eux au moins. En effet, il s'agit en fait d'individualiser les groupes d'attributs qui sont linéairement associés en vertu du jeu d'une autre dimension, latente et non directement observable. L'analyse d'une série de matrices factorielles issues des différentes trames individualisées au départ et qui prenaient en compte quelque 170 attributs spatiaux initiaux, devait ainsi révéler l'existence de toute une série de variables qui à elles seules déterminaient l'existence d'un facteur, ce qui signifiait en d'autres termes que toutes les autres saturations étaient faibles ou voisines de zéro.

La structure du facteur (ou de la composante) pouvait être alors de trois types : composante strictement unipolaire, composante de tendance unipolaire, avec sur le pôle le plus faible des attributs sans corrélation simple importante, composante définie par deux attributs de signes opposés mais entretenant entre elles une très forte corrélation négative : ainsi par exemple les pourcentages respectifs des locataires et des propriétaires. À l'exception des deux derniers, de tels attributs ne présentaient, du moins dans le cadre de notre système de recherche, aucune corrélation linéaire significative avec les autres attributs. Nous concevons fort bien cependant que la notion de « valeur significative », telle qu'employée ici peut prêter à confusion. Ayant utilisé le modèle de Hotelling, qui ne réclame pas la multinormalité des distributions, les tests de signification statistique, définis par référence à la courbe de Laplace-Gauss et en termes de probabilités, ne pouvaient nous être d'une grande utilité. Nous avons donc choisi comme *seuil* de signification une réalité plus empirique, qui était de surcroît fort saine au plan conceptuel et sans danger au plan des probabilités : un seuil oscillant entre 0,65 et 0,71 selon que les variables entretenaient plusieurs ou une seule colinéarité avec les autres ou un des autres attributs retenus au départ. Pour notre échantillon de 21 unités d'observations, soient 19 degrés de liberté, nos corrélations ont de très grandes chances d'être hautement significatives

puisque pour une distribution normale, selon la table de Fisher-Student elles le sont à 0,01 dès qu'elles atteignent le seuil de $r = 0,55$. Nous avons pourtant choisi d'élever ce seuil, nos distributions n'étant pas toutes normalement distribuées, pour nous approcher de la valeur du R^2 qui indique que la covariation des attributs se réalise dans au moins 50% des cas, ce qui géographiquement nous paraît beaucoup plus important, en tous cas tant que l'on en reste au niveau d'une analyse strictement descriptive, ne faisant pas appel à l'inférence statistique.

C'est ainsi que nous avons écarté 55 variables qui pour n'être pas inutiles dans la description d'un paysage géographique et de ses composantes ne se conformaient pas à l'optique de notre recherche, qui était de découvrir les faits les plus généraux et non les spécificités. Ces dernières devraient au besoin faire l'objet d'un autre type d'analyse, voire du même type d'analyse mais de façon indépendante si l'on découvre qu'elles font en fait partie d'un système ou d'un sous-système indépendant de celui qui a été défini par l'ensemble des attributs spatiaux retenus au départ. En matière d'utilisation du sol par exemple il devint vite évident que quatre attributs spatiaux mériteraient d'être analysés à l'intérieur de sous-systèmes particuliers dont ne pouvait rendre compte une analyse réalisée à l'échelle d'un morceau de la banlieue et à partir de données statistiques relevées au niveau des municipalités : le sous-système industriel, le sous-système institutionnel, et plus visiblement encore le sous-système du transport et de la circulation et le sous-système des parcs et des terrains de jeux.

Est-il encore nécessaire de légitimer une telle série d'éliminations au cours de différentes analyses factorielles successives ? La pertinence de notre méthode nous paraît clairement établie au triple plan de la logique mathématique, de la logique géographique et de la logique opérationnelle. Le modèle factoriel, tel qu'il est utilisé jusqu'ici est un modèle de *recherche* et non un modèle de *simulation*. L'analyse factorielle n'est rien d'autre en fait qu'un procédé d'induction quantitative, un « tamis » ou un « filtre » d'une réalité trop complexe pour être appréhendée de façon directe par l'observateur et qui, en « l'absence de fondement théoriques solides » comme le rappelle fort justement Phil Rees ¹⁸, sans doute à la suite des remarques de J. S. Armstrong ¹⁹, devient un instrument de travail permettant de faire un *tri* parmi les faits d'observation empirique avant même de songer à les utiliser pour calculer et prévoir les faits inconnus. Peut-être n'est-il pas inutile pourtant d'attirer l'attention du lecteur sur l'importance d'un autre type d'argument, lié à la nature du procédé VARIMAX généralement utilisé par la grande majorité des chercheurs. En permettant d'ajuster le système des « dimensions latentes » à une série de groupes d'attributs très distincts dont les membres ont avec la dimension des saturations maxima et sont unis par des relations très fortes, la rotation des axes minimise inversement

¹⁸ REES, Ph. Factorial Ecology, *Opus cit.*, note 2.

¹⁹ ARMSTRONG, J.S. Derivations of Theory by Means of Factor Analysis or Tom SWIFT and his Electric Analysis Machine. *The Amer. Statistician*, XXI (5) : 17-21.

les saturations obtenues par les attributs ne faisant pas partie de ces groupes. Ces saturations sont voisines de zéro. Alors que dans une solution n'ayant pas subi de rotation, les poids locaux d'une composante sont conditionnés par la présence parfois « pesante », tout en n'étant pas déterminante, de toutes les saturations, le poids de cette présence est minima lorsqu'a été appliqué le procédé VARIMAX. Le niveau de conformité existant entre la définition verbale du facteur ou de la composante et les valeurs des poids locaux s'en trouve donc automatiquement élevé.

Les dimensions latentes issues d'une solution VARIMAX s'ajustant très exactement à des groupes cohérents d'attributs, il s'en suit que le retrait d'un ou de plusieurs attributs n'affecte pas ou très peu la nature des groupements obtenus et donc la définition de ces dimensions, facteurs ou composantes, tout en permettant d'éliminer le ou les attributs dont les valeurs ont entraîné lors d'une première réduction factorielle des distorsions dans la valeur et le classement des poids locaux.

L'inverse semble aussi vrai. L'utilisation « intelligente » de la solution VARIMAX suppose à notre avis qu'au départ la matrice soit nourrie de relations assez fortes au sein des différents groupes d'attributs. Dans le cas contraire la solution n'est guère efficace. Elle aboutit en fait à détecter les attributs qui n'entretiennent pas avec d'autres de corrélations intéressantes, ces attributs étant en quelque sorte « rejetés » sur des axes moins importants formant autant de dimensions qui s'apparentent presque à des « facteurs spécifiques », chacune étant indexée par le seul attribut qui a obtenu une saturation assez forte pour qu'on l'utilise dans la définition verbale du facteur ou de la composante. Les poids locaux sont alors difficilement utilisables, tant au plan conceptuel qu'au plan algébrique, puisqu'ils sont malgré tout « nourris » par la totalité des rapports, même très faiblement exprimés au plan des saturations, unissant chacun des attributs considérés au départ à la dimension latente considérée. Il nous paraît donc évident que la nature même du procédé de rotation VARIMAX, qui recherche une structure factorielle simplifiée, conduit par analogie le chercheur à se débarrasser de tous ces attributs encombrants qui nuisent à la conformité qui doit toujours exister entre les résultats obtenus dans le simulacre mathématique et ceux que l'on peut observer dans la réalité empirique, celle des faits géographiques.

d) Le test de conformité factorielle

L'analyse des différentes trames ainsi « épurées » devait alors donner des résultats dont on pouvait tester la conformité avec la réalité empirique, telle que traduite par les valeurs des attributs originaux « entrés » dans les différentes matrices d'information. À cette fin et pour chacune des trames, nous avons confronté à chacun des attributs spatiaux ayant été retenus après une première sélection dans la définition des composantes structurant chacune des trames, chacun des poids locaux obtenus par chacune des unités d'observation sur chacune de ces composantes. En choisissant des critères

Tableau 1 *Tableau testant la conformité des poids locaux de la trame des commerces par rapport aux valeurs standardisées des attributs qui l'expliquent.*

NOM DES MUNICIPALITÉS	Variété des commerces		Magasins rares		Vêtements et accessoires		Équipement domiciliaire		Alimentation spécialisée		Nombre de magasins		Magasins rares (indice)		Activité de l'automobile		Marchandises généralistes			
Longueuil	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	(X)		Poids local très élevé	
Jacques-Cartier Saint-Lambert	×	×	×	×	×	×	×	×	×	×	×	×	(-)	×	(-)	×	×	×	×	Poids local élevé
Greenfield Park Chateauguay-Centre Laprairie Boucherville	×	×	(X)	(X)	×	×	×	×	(-)	(-)	(X)	(X)	×	×	(X)	×	×	×	×	Poids local relativement élevé
Saint-Hubert Saint-Bruno Lafèche	(X)	(X)	(-)	(-)	(X)	(X)	×	(-)	(X)	(-)	(X)	(-)	(-)	×	×	×	×	×	Poids local moyen (Positif)	
Chateauguay Lemoyne Brossard	(-)	(-)	(-)	(-)	(X)	(-)	(-)	(-)	(X)	(-)	(-)	(-)	×	×	(-)	(X)	(X)	(=)	Poids local moyen (Négatif)	
Saint-Constant Candiac Delson Notre-Dame	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(X)	(X)	(X)	(=)	(=)	(X)	Poids local relativement faible	
Chateauguay-Height Léry Sainte-Catherine Prévile	(=)	(=)	(=)	(=)	(=)	(=)	(=)	(=)	(-)	(-)	(=)	(=)	(=)	(=)	(=)	(=)	(=)	(=)	Poids local faible	

Légende : × × Supérieure à 1 écart-type
 × Entre 0,5 et 1 écart-type
 (X) Entre 0 et 0,5 écart-type
 (-) Entre 0 et -0,5 écart-type
 - Entre -0,5 et -1 écart-type
 = Supérieur à -1 écart-type

et des valeurs-seuils bien définis, on pouvait faire faire ce travail à l'ordinateur. Mais dans le cadre de ce tout premier essai, nous avons préféré travailler à la main et le tableau 1 illustre la méthode utilisée pour tester la conformité des signes, voire des valeurs existant entre les poids locaux et les attributs spatiaux. Dans la première colonne de gauche, nous avons disposé les noms des municipalités de la couronne urbaine de Montréal selon l'ordre de leur importance dans la hiérarchie, du pôle positif au pôle négatif, des poids locaux. À cette colonne des municipalités correspond toute une série de rangées permettant de lire (selon l'ordre hiérarchique des saturations sur les composantes) les valeurs originales standardisées des attributs, valeurs réduite à un ou deux signes indiquant par rapport à une moyenne zéro et un écart-type égal à l'unité, le sens et l'importance de leur rôle dans la définition de la composante. Les signes x ou $-$ ou xx et $=$ indiquent en effet la position de l'observation par rapport à la valeur zéro. La présence de deux signes côte à côte signifie que l'écart à la moyenne a atteint ou dépassé le seuil d'un écart-type.

Cette méthode nous a permis d'obtenir pour chacune des composantes une vue d'ensemble du classement factoriel en regard des valeurs originales standardisées des attributs spatiaux. Nous avons pu immédiatement détecter les municipalités qui étaient sous ou sur-évaluées par les poids locaux en regard des valeurs qu'elles obtenaient sur les attributs ayant généré la composante et la série de ses poids locaux. Le tableau 2 donne le résultat global de ce test, appliqué à 25 composantes obtenues à partir des neuf trames analysées. Sur un total de 525 poids locaux nous n'avons détecté qu'un faible 6,9% de « non conformité ». Dans la plupart des cas l'erreur n'avait guère d'importance d'ailleurs et était toujours explicable. Avec les valeurs limites choisies, fixes et finalement très empiriques, il ne pouvait en être autrement. C'est dire que notre méthode n'est pas définitive, mais c'est dire aussi que tel quel, notre test permet manifestement de confirmer la sécurité de notre méthode de traitement factoriel. Identifiant les poids locaux « anormaux » l'information recueillie par ce test ajoute en outre un renseignement de plus à leur « pedigree » et permet le cas échéant de comprendre dans les étapes subséquentes de l'analyse, les anomalies liées à des valeurs contestables.

En définitive, ce choix des attributs spatiaux significatifs au niveau du système spatial considéré à partir du double examen des communautés et des linéarités associant deux à deux chacun de ces attributs, introduit la notion d'*itérations factorielles*, ou d'analyses factorielles de *rangs* différents, conduisant à la définition de matrices d'information plus ou moins optimisées sur lesquelles porteront les analyses finales dont on utilisera par la suite les résultats.

4. Des itérations factorielles au choix des indicateurs géographiques

Le fait de ne plus se contenter des résultats d'une première analyse factorielle mais de les utiliser pour procéder à d'autres analyses factorielles

Tableau 2 *Tableau des municipalités dont les correspondances entre les poids locaux et les attributs spatiaux sont inadéquates.*

TRAME FACTORIELLE ET COMPOSANTES	Utilisation du sol			Habitat			Statut fonctionno- professionnel		Classes d'âge			Statut familial				Statut socio- économique				Commerce		Services			Finances municipales																
	I	II	III	I	II	III	I	II	I	II	III	I	II	III	IV	I	II	III	IV	I	II	I	I	II	III																
MUNICIPALITÉS																																									
1-Boucherville	=											+	+																												
2-Brossard															++																										
3-Candiac													+	-																											
4-Chateauguay														+	-																										
5-Chateauguay-Centre			-				++							+																											
6-Chateauguay-Height																			+																						
7-Delson																																									
8-Greenfield Park			-																																						
9-Jacques-Cartier																																									
10-Lafleche																																									
11-Laprairie													+																												
12-Lemoyne															++	-																									
13-Léry							-																																		
14-Longueuil																	++																								
15-Notre-Dame																																									
16-Prévilles																																									
17-Saint-Bruno-de-Montarville																																									
18-Ste-Catherine-d'Alexandrie																																									
19-Saint-Constant													++																												
20-Saint-Hubert																																									
21-Saint-Lambert																																									

Soit, sur 525 poids locaux (21 25) 36 cas inadéquats (6,9%) dont 9 (1,7%) vraiment importants.

Légende

- ++ Municipalité dont le poids local est très sur-élevé par rapport aux valeurs positives du continuum.
- + Municipalité dont le poids local est sur-évalué par rapport aux valeurs positives du continuum.
- = Municipalité dont le poids local est sous-évalué par rapport aux valeurs positives du continuum.
- Municipalité dont le poids local est très sous-évalué par rapport aux valeurs positives du continuum.

portant sur les mêmes observations, assorties d'un nombre plus réduit de variables, correspond donc à effectuer une analyse factorielle de deuxième rang et au besoin de troisième rang.

a) *Les analyses factorielles de deuxième et de troisième rangs*

Il s'agit en effet de substituer à l'analyse originale de rang $n \times m$, l'analyse d'une matrice plus réduite, de rang $n \times m'$ dans laquelle m' représente le nombre d'attributs ayant résisté au filtre de la première analyse. Le nombre de facteurs d'une analyse de deuxième rang est évidemment plus réduit et leurs poids locaux sont d'une manière très générale bien plus expressifs de la réalité géographique qualifiée par les quantités retenues dans cette matrice d'information spatiale de deuxième rang. L'importance des « bruits de fonds » tend en effet à être inversement proportionnelle au nombre des saturations possibles qui entrent en compte dans le calcul de ces poids locaux. Ces bruits de fonds ne disparaissent pourtant pas complètement et il peut être parfois intéressant de réaliser une *analyse factorielle de 3ème rang* qui ne prendrait en considération que les attributs spatiaux ayant obtenu sur l'un ou l'autre des facteurs que l'on veut cartographier ou dont on veut utiliser les poids locaux pour des analyses subséquentes, de très fortes saturations, celles qui justement ont permis d'identifier le facteur. Ainsi par exemple, notre analyse factorielle de la structure des commerces de détail ou des services commerciaux et professionnels nous permettait de découvrir dans les deux cas l'existence d'un premier facteur, d'importance considérable, identifiant les types d'activité liés à la *centralité*. Ces facteurs regroupant sept attributs ayant de fortes saturations, parmi les neuf que comptait la matrice d'information il était intéressant d'isoler ce facteur de centralité en retirant les attributs ne jouant qu'un rôle très faible. Nous fûmes ainsi conduits à refaire l'analyse factorielle d'une matrice d'information spatiale encore plus réduite, analyse de troisième rang, la matrice soumise à l'algorithme n'étant nourrie que des seuls attributs ayant obtenu sur le facteur dit de « centralité » des saturations dépassant un certain seuil. L'analyse ne donne alors qu'un seul facteur, ou une seule composante, et bien évidemment qu'une seule solution, la rotation d'axe n'ayant dans ce cas aucun sens.

Le seuil retenu dans le premier cas nous a été dicté par une très nette coupure entre les variables obtenant une saturation de 0,80, toutes les autres se situant à un niveau nettement plus bas : 0,50 et moins. Pour les « services » le seuil fut choisi au niveau de la saturation de 0,67, toutes les autres variables obtenant des saturations faiblement négatives. Mais encore une fois l'analyse des communautés doit confirmer la décision en cette matière tout comme la prise en considération des colinéarités significatives, en regard du seuil limite, toujours le même, celui du R^2 égal ou supérieur à 50%, soit une corrélation linéaire (ou une saturation) supérieure à 0,71. Ce type d'analyse a cependant un inconvénient : lorsque le facteur est bipolaire, l'analyse de troisième rang donnera évidemment non pas un seul

facteur mais deux facteurs unipolaires, indépendants l'un et l'autre, pour lesquels les poids locaux, de nouveau, seront marqués par la présence, même extrêmement faible, dans le processus de calcul, des saturations obtenues par les attributs indexant l'autre facteur. Dans ce cas il vaut mieux ne retenir que les variables (positives ou négatives) dont on veut absolument tirer un indicateur composé et synthétique n'exprimant rien d'autre que la dimension latente unissant les seuls attributs considérés et procéder à deux analyses factorielles ne donnant chacune qu'un facteur ou une seule composante principale.

On se rappellera cependant qu'un paysage est un tout, de même que le « système » qui le formalise en vue de l'analyse statistique. Il est donc nécessaire, pour en comprendre l'agencement et faire écho à sa réalité profonde, de le saisir dans son ensemble et non plus à partir de chacune des trames qui le composent. L'analyse factorielle des trames et leur réduction en un certain nombre de dimensions indépendantes ne peut être une fin en soi. Nous pouvons, à travers une série d'algorithmes de classification réduire chacune de nos trames à une typologie optimale des observations et comparer ces typologies les unes aux autres, soit par la superposition et l'analyse de neuf cartes différentes soit encore utiliser comme J. B. Racine et Guy Lemay²⁰ l'ont proposé, et illustré, l'algorithme des *correspondances discriminatoires dans l'espace géographique*. La méthode permet en effet de relier les différentes trames typologiques issues de différentes analyses factorielles et de déterminer leur niveau de correspondance à l'aide d'un coefficient qui s'apparente au R^2 tout en n'étant pas réversible comme l'est le carré d'un coefficient de corrélation simple. La méthode peut se révéler extrêmement utile au plan de l'analyse géographique et de l'interprétation des faits géographiques, de la différenciation et de leur organisation. Ici cependant nous avons pensé que la nature des poids locaux « purgés des bruits de fonds » mathématiques qu'ils contenaient nous permettait de les réutiliser dans un nouveau modèle factoriel pour lequel la matrice d'information serait nourrie par les indicateurs factoriels « composés » que nous venions de découvrir pour chacune de nos trames. Ces indicateurs composés, sous forme de poids locaux, résumaient en effet le plus adéquatement possible l'ensemble de l'information originale. De nouveau cependant un choix préalable était nécessaire.

b) *Le choix des indicateurs composés*

Un seul problème se posait : fallait-il introduire tous les indicateurs composés, toutes les séries de poids locaux révélées par les différentes trames étudiées ou n'en choisir qu'un certain nombre ? Une première élimination avait de soi : deux composantes de nature plus ou moins analogue ne devaient pas être prises en compte simultanément. Ainsi par exemple on pouvait obtenir une dimension appelée « type d'habitat » aussi bien à partir

²⁰ RACINE, J.B. & LEMAY, Guy (1972), L'analyse discriminatoire des correspondances typologiques dans l'espace géographique. *L'espace géographique*, 1 (3).

de l'analyse de la trame d'utilisation du sol, de la trame du logement et même de la trame du statut familial, vu que dans cette dernière nous avons introduit des indicateurs identifiant la nature du logement. Le choix du meilleur résultat en termes de correspondance entre la définition du facteur et la hiérarchie des poids locaux s'imposait à l'évidence. Pour le reste il suffisait d'appliquer à une première analyse factorielle des poids locaux les mêmes principes que ceux qui ont conduit à l'épuration des trames en croisant les résultats obtenus dans la matrice des corrélations entre poids locaux et ceux donnés par le calcul des communautés. Sur les vingt-trois poids locaux mesurant la distribution des composantes de l'espace géographique de la zone métropolitaine de Montréal au sud du Saint-Laurent, une quinzaine seulement remplissaient les conditions requises.

L'analyse des résultats (Tableau 3) fut extrêmement satisfaisante tant au plan mathématique que géographique. En outre les composantes découvertes correspondaient exactement à ce que nous pressentions au niveau théorique, et à ce que nous pouvions prévoir de l'examen serré des différentes variables entrées au départ dans chacune de nos trames. On en trouvera un commentaire précis et circonstancié dans les prochains travaux à paraître : le texte de la maîtrise en géographie de Marc Cavalier²¹ et d'une façon plus générale dans le gros ouvrage que J. B. Racine²² termine actuellement sur la *croissance du Grand Montréal*. Si l'analyse géographique s'arrêtait là, il est certain que c'est sans doute ce résultat qui devrait être à la base de la réflexion géographique du chercheur. Cependant, pour celui qui veut emprunter ces résultats (les poids locaux en particulier) pour des analyses subséquentes (algorithmes de classification ou essais de modélisation), plusieurs problèmes peuvent alors se poser. D'abord celui du « degré de fraîcheur des données » : il deviendrait vite minimal ou difficile à contrôler. L'interprétation des résultats d'analyses subséquentes serait en outre extrêmement complexe et délicate puisqu'il conviendrait à chaque étape d'avoir recours à « l'arbre généalogique » des facteurs et au « pedigree » des attributs composés utilisés, et de réexpliquer la nature des valeurs découvertes en fonction de leur création mathématique originelle. Surtout, et c'est peut-être là l'essentiel, ces données ne sont pas directement disponibles pour d'autres chercheurs, voire pour le même chercheur, voulant tester la généralité de son résultat en s'appliquant à l'étude d'autres régions ou d'autres milieux. Ne pouvant jamais faire partie d'une banque générale de données officielles, elles obligent chacun des analystes à procéder lui-même à la même démarche que nous avons suivie pour parvenir aux indicateurs dont il aurait besoin, ce qui lui demandera du temps et toute une série de manipulations intermédiaires fastidieuses, sans que jamais il soit assuré d'arriver aux mêmes résultats.

²¹ CAVALIER, Marc (1972), *Regroupement municipal et indicateurs géographiques*. Thèse de Maîtrise, Université d'Ottawa.

²² RACINE, J.B. *Un type nord-américain d'expansion métropolitaine : la couronne urbaine du Grand Montréal (géographie factorielle expérimentale d'un phénomène suburbain)*. Ottawa, Éditions de l'Université d'Ottawa (en préparation).

Tableau 3 *Analyse des indicateurs factoriels composés de la zone métropolitaine de Montréal au sud du Saint-Laurent (Solution Varimax)*A— *Résultats généraux* : 4 composantes ; poids cumulé de 82,10%

<i>Identification des composantes</i>	<i>Valeurs propres</i>	<i>Poids</i>
C I : Statut socio-économique	4,24	28,29%
C II : Centralité économique	2,91	19,38%
C III : Intensité de l'occupation urbaine	2,65	17,70%
C IV : Statut familial	2,51	16,72%

Communautés (dans l'ordre des attributs)

0,70 - 0,90 - 0,62 - 0,80 - 0,93 - 0,74 - 0,94 - 0,87 - 0,89 - 0,82 - 0,73 - 0,98 - 0,88 - 0,89 - 0,62.

B— *Matrice factorielle*

<i>Attributs spatiaux</i>	<i>Composantes et saturations</i>			
	I	II	III	IV
1- Faiblesse de l'habitat collectif	0,02	-0,22	0,80	0,06
2- Importance de l'habitat unifamilial	-0,23	-0,46	<u>0,77</u>	-0,19
3- Médiocrité de la qualité du logement	<u>0,70</u>	0,05	0,33	-0,15
4- Importance des adolescents	<u>0,38</u>	-0,15	-0,77	-0,18
5- Maturité démographique	-0,80	-0,15	0,40	0,31
6- Vieillesse démographique	<u>0,27</u>	0,14	-0,04	<u>0,81</u>
7- Famille-type de la banlieue	0,02	0,20	-0,11	-0,94
8- Familles nombreuses	<u>0,81</u>	0,02	0,46	0,00
9- Médiocrité socio-économique	<u>0,91</u>	0,02	-0,12	0,20
10- Faiblesse des emplois féminins	<u>0,23</u>	-0,05	0,31	-0,82
11- Importance des professions secondaires	<u>0,84</u>	-0,13	-0,08	-0,02
12- Centralité commerciale	-0,09	<u>0,94</u>	-0,29	0,08
13- Centralité des services	-0,24	<u>0,86</u>	-0,25	-0,15
14- Taille et richesse municipales	<u>0,11</u>	<u>0,93</u>	-0,08	0,01
15- Équilibre de la gestion	-0,72	<u>0,20</u>	-0,20	-0,18

C— *Matrice des poids locaux des composantes*

<i>Unités d'observations</i>	<i>Composantes et poids locaux</i>			
	I	II	III	IV
1- Boucherville	-0,104	0,686	0,084	-1,259
2- Brossard	0,055	-0,454	-0,480	-0,767
3- Candiac	-0,822	-0,677	0,168	-1,568
4- Chateauguay	-0,366	-0,142	-0,437	-0,173
5- Chateauguay-Centre	-0,226	0,817	0,799	-0,558
6- Chateauguay-Height	-0,624	-1,413	0,344	2,931
7- Delson	0,752	-1,110	0,160	0,106
8- Greenfield Park	-0,674	0,437	-0,825	0,468
9- Jacques-Cartier	1,302	2,423	0,853	0,215
10- Laflèche	0,982	0,072	0,106	0,618
11- Laprairie	0,448	0,211	-0,963	0,455
12- Lemoyne	0,476	-0,759	-2,836	-0,553
13- Léry	-0,103	-0,903	0,229	0,937
14- Longueuil	-0,282	1,459	-1,762	0,645
15- Notre-Dame	1,307	-0,371	1,477	0,295
16- Prévile	-2,126	-1,342	1,278	-0,689
17- Saint-Bruno-de-Montarville	-1,003	0,304	0,380	-1,330
18- Sainte-Catherine-d'Alexandrie	1,391	-0,859	0,271	-0,723
19- Saint-Constant	1,407	-0,698	0,149	-0,136
20- Saint-Hubert	0,224	0,794	0,955	0,112
21- Saint-Lambert	-1,991	1,453	-0,007	1,039

C'est pourquoi, en faisant provisoirement l'hypothèse que les quatre dimensions découvertes étaient des dimensions d'une grande généralité *, nous nous sommes demandés s'il ne convenait pas plutôt de chercher à les représenter par des indicateurs simples d'accès universel et qui rendraient compte des mêmes dimensions tant sur le plan géographique (spatial) que sur le plan conceptuel.

c) *Le choix d'indicateurs simples*

On doit se demander en effet si un attribut ayant obtenu la plus forte des saturations (positive ou négative) sur l'un ou l'autre des facteurs issus d'une analyse, de n'importe quel rang d'ailleurs, n'est pas d'un emploi plus sûr et plus facile pour les analyses subséquentes, que ces dernières soient strictement cartographiques ou encore une fois d'ordre statistique. La géographie volontaire ou « appliquée » doit forcément, pour être efficace, se nourrir d'attributs simples et d'accès aisé. Des tests de corrélations entre la distribution des attributs arrivant en tête des pôles d'une composante et la distribution des poids locaux obtenus par cette même composante nous ont convaincus que, dans la majeure partie des cas, le choix d'un indicateur simple, dont on connaît cependant le « pedigree » factoriel et par là le rôle dans le système analysé, est de ce point de vue une excellente solution, parfois la meilleure possible de toutes manières. La description géographique y perd généralement moins de 10% de la complexité enregistrée par la dimension latente, mais ces 10% sont largement compensés au plan scientifique, technique et opérationnel par la sécurité et la simplicité liées au recours à des indicateurs que tout le monde peut comprendre et auxquels chacun peut avoir accès. C'est ainsi par exemple que pour le facteur dit de « centralité », les variables indexant soit la variété des types de commerces soit la variété des types de services commerciaux traduisent aussi bien le phénomène qu'un résultat issu d'une analyse factorielle de 3^{ème} rang.

On ne saurait toutefois se contenter systématiquement des attributs obtenant les plus fortes saturations sur les facteurs. Ici la réflexion théorique reprend ses droits. En effet, si un indicateur peut *mathématiquement* représenter plusieurs autres attributs, il ne peut pas toujours le faire sur le plan *conceptuel*. Deux variables peuvent exprimer exactement la même réalité géographique : ainsi par exemple les ouvriers et les manœuvres, unis par une très forte corrélation linéaire et se rangeant l'une à côté de l'autre sur le même pôle de la composante. Elles peuvent très bien être prises l'une

* On aura certainement remarqué l'absence du *facteur ethnique* dans notre résultat. Comme on le verra plus loin, et conformément au problème fondamental de la métropole du Québec, nous n'avons introduit dans l'analyse qu'une seule donnée ethnique : le « taux de francité », autrement dit le « non-french » de l'analyse de GREER-WOTTEN ²³. Dans tous les cas cette variable s'est greffée à la composante du statut socio-économique, sur le pôle qui indexait sa médiocrité.

²³ GREER-WOOTTEN, Bryn (1972), Changing Social Areas and the Intra-Urban Migration Process. *Rev. Géogr. Montr.*, XXVI (3): 271-292.

pour l'autre, l'une représentant l'autre de la même manière que l'importance des propriétaires dans une municipalité traduit automatiquement la faiblesse proportionnelle des locataires. De nouveau l'une ou l'autre de ces variables, même si elles se situent pourtant aux deux pôles opposés de la même composante, parce qu'elles entretiennent, selon la façon de les mesurer, une corrélation négative voisine ou égale à l'unité, peut être seule prise en compte, étant entendu que la présence de l'une implique l'absence de l'autre et inversement. Mais en revanche deux variables peuvent être en tête d'un facteur sans exprimer automatiquement la même réalité : si à Montréal le « taux de francité » covarie avec l'importance des ouvriers, des manœuvres et du taux de chômage et se retrouve sur le même pôle (celui de la médiocrité) du facteur identifiant le statut socio-économique, les réalités géographiques que chacune d'elles exprime individuellement ne sont pas les mêmes. Ces variables sont complémentaires et il faut absolument les conserver toutes les trois pour rendre compte de la réalité géographique synthétique particulière à la région analysée, réalité que l'on n'entrevoit au départ que par hypothèses liées soit à une première série d'observations, soit aux travaux déjà effectués sur la région ou sur le problème considéré.

C'est en suivant l'ensemble de ces règles que nous sommes finalement parvenus, à partir des résultats de l'analyse des indicateurs composés, à identifier la série d'attributs importants dont nous avons besoin pour notre analyse finale. Mais parallèlement nous devons constituer une matrice de corrélation réduite unissant deux à deux chacun des attributs simples susceptibles d'être retenus, de manière à vérifier que deux variables de nature plus ou moins analogue ne soient pas prises en compte simultanément. Nous devons ainsi découvrir, ce qui s'avéra fort utile, l'importance des corrélations unissant les différentes classes d'âge telles que traditionnellement utilisées dans les pyramides d'âge par rapport à celles identifiant l'âge des chefs de famille. Ayant préalablement découvert par l'analyse factorielle de la pyramide des âges que le découpage du recensement n'était pas conforme à la structure des regroupements démographiques et ayant établi un nouveau découpage à la lumière de nos résultats²⁴, ce sont les variables les plus générales, celles des classes d'âges que nous avons reconstituées que nous avons utilisées. Encore fallait-il pour notre analyse finale choisir les classes les plus représentatives de l'ensemble, en vérifiant les résultats de notre analyse factorielle de second rang et en les resituant dans le contexte global de l'étude. On évitait ainsi les dédoublements superflus tout en choisissant les attributs ayant manifestement avec les indicateurs d'autre nature les plus fortes corrélations. C'est ainsi qu'ont été choisis cinq indicateurs d'âge : les 5-9 ans, les 25-29 ans, les 30-39 ans, les 40-49 ans et les 50-64 ans, le « troisième âge », celui de la retraite, ne jouant en banlieue qu'un rôle extrêmement faible et étant de toute façon en étroite corrélation avec la classe d'âge qui le précède, celle des 50-64 ans : 0,91.

²⁴ RACINE, J.B. (1972), Voir note 1.

On comprend immédiatement ici que le nombre des attributs simples sera forcément supérieur au nombre des attributs composés. De trois attributs démographiques on est obligé de passer à cinq. Mais le nombre total des indicateurs simples finalement retenus n'est que d'une vingtaine. En fait quatre composantes ne sont pas directement représentées par l'un des attributs qui l'indexait. La composante « importance des emplois secondaires » qui exprimait l'opposition « secondaire-tertiaire » était indexée par des attributs moins représentatifs que les « ouvriers », les « artisans » et les « administrateurs » contenus dans la composante « la médiocrité du statut socio-économique ». Quant aux composantes « médiocrité de la qualité du logement » et « équilibre de la gestion municipale », elles n'obtenaient pas les communautés réclamées dans le premier filtrage factoriel des poids locaux et elles obligeaient à ajouter trois variables supplémentaires qui seraient venues se greffer, comme nous l'indiquait la matrice des corrélations simples, sur un « statut socio-économique » déjà indexé par six variables. Le déséquilibre était parfaitement prévisible. Cette élimination entraînait « ipso facto » celle de l'attribut « familles nombreuses » qui tout en entretenant avec deux de ces derniers attributs éliminés de fortes colinéarités était déjà représenté par ses corrélations avec les variables indexant la médiocrité du statut socio-économique. Dernier changement enfin, pour réduire le nombre des attributs au minimum, la substitution de la composante « habitat collectif » par un seul attribut, la densité résidentielle, qui contient implicitement dans ses variations les variations de la densité et du type d'habitat. Le tableau 4 définit en même temps que la structure factorielle de la zone métropolitaine de Montréal au sud du Saint-Laurent, les indicateurs géographiques finalement retenus. Ces résultats devaient nous permettre alors de passer vraiment des unités statistiques quantitatives aux unités géographiques qualitatives.

Nous gardions cependant en réserve une troisième série de résultats, *mixtes*, nés d'une matrice d'information spatiale unissant des indicateurs simples et des indicateurs composés pouvant éventuellement mieux représenter certaines situations géographiques complexes tels par exemple les poids locaux de la composante indexant le statut socio-économique ou celle du statut familial, voire de la centralité des services, autant de résultats factoriels issus d'analyses de deuxième ou de troisième rang. Dans ces derniers cas en effet, contrairement à ce qui se passait pour des indicateurs simples telles que « maisons unifamiliales », « moyenne d'enfants par famille », « participation des femmes à la main d'œuvre » ou « densité résidentielle », qui s'imposaient non seulement dans les termes de notre méthode mais par évidence logique élémentaire, nous pouvions avoir quelques doutes quant à la représentativité des attributs simples. Nous verrons plus loin que cette précaution ne devait pas s'avérer inutile.

II. DES ALGORITHMES INFORMATIQUES À L'ANALYSE GÉOGRAPHIQUE

La première étape consistait donc à obtenir en définitive trois matrices factorielles à partir de trois matrices d'informations spatiales réduites, con-

Tableau 4 *Analyse des indicateurs factoriels simples de la zone métropolitaine au sud du Saint-Laurent (Solution Varimax)*A— *Résultats généraux* : 4 composantes ; poids cumulé de 86,41%

<i>Identification des composantes</i>	<i>Valeurs propres</i>	<i>Poids</i>
C I : Statut socio-économique	5,49	27,45%
C II : Statut familial	4,65	23,27%
C III : Intensité de l'occupation urbaine	4,22	21,12%
C IV : Centralité économique	2,91	14,58%

Communautés

0,86 - 0,88 - 0,72 - 0,92 - 0,95 - 0,96 - 0,84 - 0,97 - 0,74 - 0,93 - 0,85 - 0,72 - 0,95 - 0,90 - 0,79 - 0,81 - 0,92 - 0,84 - 0,92 - 0,80.

B— *Matrice factorielle**Composantes et saturations*

<i>Attributs spatiaux</i>	I	II	III	IV
1— Densité résidentielle	0,04	-0,15	<u>0,91</u>	0,13
2— Trois enfants (%)	-0,07	<u>0,89</u>	<u>0,03</u>	0,27
3— Moyenne d'enfants/famille	0,27	<u>0,79</u>	0,07	0,12
4— Propriétaires (%)	0,39	- <u>0,26</u>	- <u>0,82</u>	0,16
5— Maisons unifamiliales (%)	-0,26	0,24	- <u>0,85</u>	-0,32
6— Autres appartements (%)	0,17	-0,16	<u>0,92</u>	0,26
7— Participation des femmes à la main-d'œuvre (%)	-0,07	- <u>0,84</u>	<u>0,25</u>	0,25
8— Salaire moyen (\$)	- <u>0,98</u>	- <u>0,02</u>	-0,07	-0,11
9— Taux de francité	<u>0,71</u>	0,35	0,34	-0,04
10— Administrateurs (%)	- <u>0,94</u>	0,02	-0,09	-0,18
11— Ouvriers et artisans (%)	<u>0,92</u>	0,03	0,11	-0,01
12— Indice de chômage	<u>0,82</u>	-0,10	0,20	-0,05
13— Variété des commerces	<u>0,08</u>	-0,07	0,35	<u>0,90</u>
14— Classe d'âge 5 - 9 ans (%)	0,07	0,75	- <u>0,55</u>	-0,15
15— Classe d'âge 25 - 29 ans (%)	0,60	0,36	<u>0,54</u>	0,12
16— Classe d'âge 30 - 39 ans (%)	-0,13	<u>0,82</u>	- <u>0,34</u>	0,08
17— Classe d'âge 40 - 49 ans (%)	- <u>0,91</u>	- <u>0,28</u>	-0,04	-0,04
18— Classe d'âge 50 - 64 ans (%)	- <u>0,25</u>	- <u>0,86</u>	0,17	0,07
19— Variété des services	0,14	<u>0,14</u>	0,19	<u>0,92</u>
20— Valeur des biens-fonds	-0,03	0,01	0,12	<u>0,88</u>

C— *Matrice des poids locaux des composantes**Composantes et poids locaux*

<i>Unités d'observations</i>	I	II	III	IV
1— Boucherville	-0,132	1,394	0,228	0,161
2— Brossard	0,410	0,516	-0,649	0,282
3— Candiac	-0,559	1,569	-0,620	-0,445
4— Chateauguay	-0,122	0,332	-0,869	0,503
5— Chateauguay-Centre	0,030	0,515	-1,012	1,050
6— Chateauguay-Height	-0,400	-3,051	-1,664	-1,021
7— Delson	0,813	-0,002	-0,313	-0,889
8— Greenfield Park	-0,654	-0,359	0,069	1,123
9— Jacques-Cartier	0,964	-0,035	0,487	1,540
10— Laflèche	0,962	-0,447	0,028	0,036
11— Laprairie	0,702	-0,393	0,823	0,250
12— Lemoyne	0,216	0,277	3,523	-1,338
13— Léry	0,148	-0,997	-0,439	-1,033
14— Longueuil	-0,149	-0,917	1,843	1,135
15— Notre-Dame	1,093	0,238	-0,002	-1,403
16— Prévile	-2,991	0,669	0,346	-1,778
17— Saint-Bruno-de-Montarville	-1,126	1,330	-0,485	0,640
18— Sainte-Catherine-d'Alexandrie	1,093	0,863	0,402	-1,352
19— Saint-Constant	0,850	0,310	-0,301	-0,554
20— Saint-Hubert	0,286	0,245	-0,561	0,640
21— Saint-Lambert	-1,743	-1,275	0,768	1,171

tenant différentes séries d'indicateurs géographiques, « composés », « simples » ou « mixtes », analogues dans leur nature mais de types différents. Les trois séries d'analyse portent sur les 21 municipalités de la zone métropolitaine de Montréal situées au sud du Saint-Laurent, la réserve indienne de Caughnawaga ayant été exclue à la suite de l'impossibilité dans laquelle nous étions de réunir pour cette observation une documentation analogue à celle réunie pour toutes les autres.

1. *L'algorithme de classification*

Il s'agissait alors d'élaborer, à partir de ces trois séries de résultats, trois trames typologiques optimisées par les itérations discriminatoires proposées à l'origine par Emilio Casetti²⁵ et que l'on commence à bien connaître depuis la grande mise au point sur l'analyse statistique en géographie proposée en 1969 par Leslie King²⁶. La première trame, issue des poids locaux découverts par l'analyse factorielle des attributs composés serait dite la trame typologique « composée », la seconde trame, issue des poids locaux découverte par l'analyse factorielle des attributs simples prenant le nom de trame typologique « simple », la dernière, parce qu'issue à l'origine d'attributs « mixtes » devenant la trame typologique « mixte ». La comparaison des trois séries de groupements optimisés par l'analyse de leurs « correspondances typologiques » (en affectant à une structure typologique les poids locaux issus d'une autre analyse et en soumettant cette nouvelle matrice d'information aux itérations discriminatoires) pouvait servir de test final quant à l'optimalité des différents choix effectués en matière d'indicateurs géographiques : indicateurs composés, indicateurs simples, indicateurs mixtes.

Au départ, l'algorithme de classification choisi fut celui que J.B. Racine recommandait dans son récent travail sur les « algorithmes de classification », l'utilisation d'une matrice d'information transposée permettant d'obtenir des *corrélations entre observations* et non plus entre attributs. À cette matrice on pouvait en effet appliquer directement l'algorithme proposé en 1957 par McQuitty²⁷, le « linkage analysis » ou « analyse des chaînes d'associations statistiques maximales », pour découvrir la « structure typique »

²⁵ CASETTI, E. a — Multiple Discriminant Functions. *Tech. Report no. 11, Computer Application in the Earth Sciences Project*. Evanston, Dept. of Geogr., Northwestern Univ.

— b — Classificatory and Regional Analysis by Discriminant Iteration. *Computer Application in the Earth Science Project, Tech. Report no. 12*. Evanston, Dept. of Geogr., Northwestern Univ. 95 p.

²⁶ KING, L., a (1969), *Statistical Analysis in Geography*. New Jersey, Prentice Hall. 228 p.

— b (1970), Discriminant Analysis : a Review of Recent Theoretical Contributions and Applications. *Economic Geography*, 46 (2) : 367-378.

²⁷ McQUITTY, L.L. (1951), Elementary Linkage Analysis for the Isolation of Orthogonal and Oblique Types and Typal Relevancies. *Educational and Psychological Measurement*, 17 : 207-229.

de l'espace considéré. Par rapport aux précédents travaux de J.B. Racine nous devons cependant apporter ici une variante importante qui rendait l'algorithme global (qui englobe les itérations discriminatoires) plus conforme à sa propre logique, plus conforme aussi à la logique de la définition préalable du résultat cherché par le géographe. Au lieu de mettre en évidence par la méthode des chaînes d'associations statistiques maximales les groupements « contenus » dans une matrice mettant en corrélation les observations assorties de *tous leurs attributs géographiques*, c'est plutôt la matrice définissant « l'espace des attributs » au sens où l'entend généralement B.J.L. Berry, l'espace des poids locaux, que nous avons transposée, avant d'utiliser l'algorithme de McQuitty. Et c'est à ce résultat là que nous avons alors appliqué les tests des itérations discriminatoires. Le résultat obtenu était bien conforme à ce que nous voulions trouver : *les différents types de combinaisons spatiales* tels qu'on pouvait les définir à partir *des principales composantes de la différenciation spatiale*, ces composantes étant exprimées par leurs poids locaux, ceux qui ont été découverts dans chacune des trois analyses finales.

Les trois séries de poids locaux obtenus à partir de nos trois analyses factorielles finales furent donc utilisées dans l'algorithme de classification et d'optimisation qui vient d'être décrit. Évidemment les trois séries de résultats ne se recouvrent pas entièrement en dépit de l'utilisation d'un seul et même procédé de calcul. Les espaces « attributs » sont de même nature mais les valeurs découvertes sont cependant différentes et elles suffisent, puisqu'elles définissent toutes, en définitive, des êtres complexes, les poids locaux des dernières composantes mises en évidence, à faire basculer les observations mal définies dans tel ou tel groupe, au gré des valeurs caractérisant les prototypes découverts par le « linkage analysis », au gré aussi des rapports entre les centroïdes définis par l'analyse discriminatoire et ces mêmes valeurs « médianes » des poids locaux. En fait, comme dans tout processus itératif, il est préconditionné par la structure des groupes qu'on lui donne à analyser, non seulement le nombre de ces groupes (nous avons d'ailleurs fait une quantité de tests sur des structures comportant le même nombre de groupes) mais encore la disposition des observations à l'intérieur de la classification. Dans chacun des groupes, chaque observation porte des valeurs qui déterminent le centroïde ou la moyenne du groupe et par là chacune des distances existant entre chacun des éléments de la classification et chacun des centroïdes des groupes. Même si les résultats finaux sont voisins, (ils le restent d'ailleurs à partir d'un groupement fait au hasard), ils ne sont que très rarement identiques. Les itérations discriminatoires effectuent presque toujours au moins deux permutations lorsque l'on affecte à une structure typologique donnée, issue d'un quelconque algorithme de classification, les valeurs des poids locaux issus pourtant *des mêmes attributs originaux*. C'est ce que J.B. Racine a démontré dans son article sur les algorithmes de classification. Nous avons cependant à notre disposition des trames analogues dans *leur structure factorielle* quoique légèrement différentes dans la structure des *attributs composés* qui en étaient issus.

La question était de savoir dans quelle mesure, en utilisant la méthode dite des correspondances discriminatoires dans l'espace géographique, déjà mentionné, on pouvait s'assurer de la parenté des structures typologiques issues des trois trames successivement analysées et mieux encore rechercher, grâce à leur utilisation combinée non seulement la meilleure de ces trois structures typologiques, mais au delà, une nouvelle structure typologique qui optimiserait simultanément les trois séries de résultats possibles.

2. À la recherche d'une structure typologique optimale

La première étape a consisté à affecter les poids locaux de chacune des trois trames à chacune des trois typologies déjà optimisées par l'analyse discriminatoire. Ainsi par exemple on devait associer à chacune des observations de chacun des groupes issus de l'algorithme de classification optimisé appliqué aux poids locaux de l'analyse des attributs composés, les poids locaux obtenus par l'analyse factorielle des attributs mixtes, puis encore les poids locaux obtenus par l'analyse factorielle des attributs simples. Ce processus nous a donc conduit à procéder à six analyses de correspondance discriminatoire : deux analyses sur la structure issue de la trame 21×15 , deux analyses sur la structure typologique issue de la trame 21×20 (mixte), deux analyses sur la structure typologique issue de la trame 21×20 (simple). Les résultats sont éloquentes. Appliqués sur les deux autres typologies, les poids locaux de l'analyse des attributs composés (21×15) ne provoquent aucun changement, et dans les deux cas l'échange des valeurs prises en compte à l'origine par le programme des itérations discriminatoires augmente considérablement la valeur du rapport

$$R = \frac{\text{variabilité intergroupes}}{\text{variabilité intragroupes}} = \frac{p}{r} = \sum \lambda r ; 0 < R < \infty ?$$

Cette somme « Σ » des valeurs-propres « λ » (somme des « eigenvalues ») passe en effet de 6,35 à 12,76 dans le premier cas et de 6,35 à 7,70 dans le second, sans qu'intervienne la moindre permutation, ce qui signifie déjà que les structures typologiques issues des deux dernières analyses sont meilleures que celle découverte à partir de l'analyse des seuls indicateurs composés. À leur tour, les poids locaux de l'analyse des attributs mixtes, lorsqu'affectés à la structure typologique des deux autres trames se traduisent dans le premier cas (typologie « composée ») par deux permutations et une diminution de la valeur de R (de 16,90 à 14,43) et dans le second cas (typologie « simple ») par deux permutations encore mais par une extraordinaire augmentation de la valeur de R qui passe de 16,90 à 43,00. On a bien ici l'indication que la classification issue de l'analyse des indicateurs « simples » est bien supérieure non seulement à celle offerte par les indicateurs « composés » mais aussi à celle que l'on dérive des indicateurs « mixtes ».

Appliquées enfin aux structures typologiques issues des trames « composées » et « mixtes », les valeurs des poids locaux tirées de l'analyse factorielle des attributs simples se traduisent dans le premier cas par une aug-

mentation de la valeur de R (qui passe de 9,75 à 12,55), mais après trois changements et cinq itérations ce qui encore une fois enlève beaucoup de crédibilité à la classification dite « composée », d'autant plus que lors de la première itération, les permutations se sont traduites par une diminution de la valeur de R. Dans le deuxième cas le résultat est meilleur : la structure dite « mixte » enregistre cinq permutations lorsqu'on lui applique les valeurs des poids locaux « simples », mais une seule itération suffit à la stabiliser, ce que démontre une deuxième itération donnant un résultat analogue à la première et la valeur de R passe de 9,75 à 33,22.

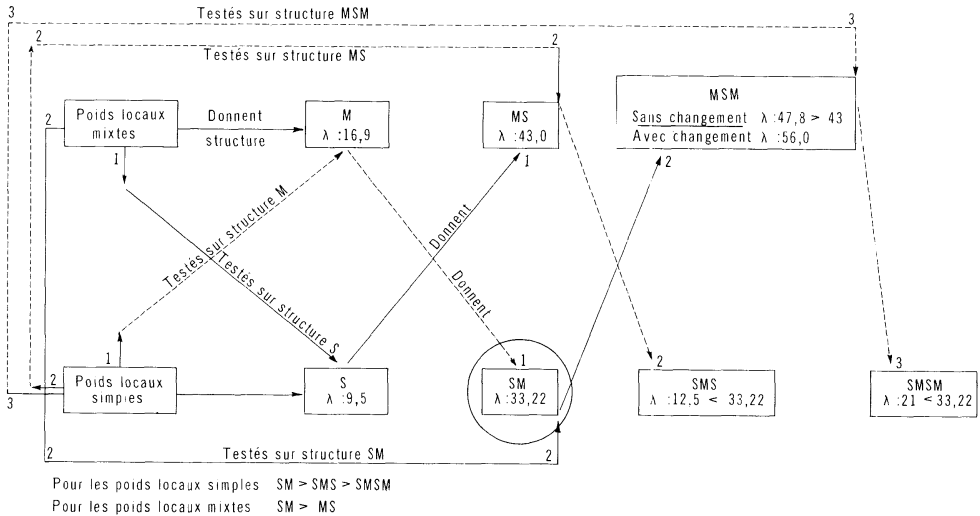
Deux résultats peuvent déjà être dégagés à cette étape du processus de comparaison des typologies. En premier lieu on doit bien constater que n'étant améliorée ni par les poids locaux de la structure factorielle mixte ni par ceux de la structure factorielle simple, la classification issue de la trame des indicateurs n'offre donc guère d'intérêt comme structure optimale, d'autant plus que ses poids locaux, testés sur les deux autres classifications démontrent que celles-ci leur conviennent même mieux encore. C'est d'ailleurs assez dire la *profonde ambiguïté* de l'analyse discriminatoire lorsque l'on veut s'en servir directement pour une recherche d'optimisation. Celle-ci sera toujours toute relative²⁸. Arrangées autrement, les mêmes valeurs donnent en effet, avec la même méthode, d'autres résultats qui peuvent être bien meilleurs. En deuxième lieu il est devenu évident qu'un choix restait à faire entre les deux autres structures typologiques, choix délicat puisque chacune d'elles était améliorée par l'application des poids locaux issus de l'analyse factorielle des « autres » indicateurs, les indicateurs mixtes dans le cas de la structure typologique issue des indicateurs simples, les indicateurs simples dans le cas de la structure typologique issue des indicateurs mixtes. On aurait pu être tenté de garder simplement la trame qui avait donné la plus haute valeur de R c'est-à-dire (cf. paragraphe précédent) la trame des indicateurs simples, telle qu'améliorée par le recours aux indicateurs mixtes puisque ce recours permet au R d'atteindre la valeur de 43,00. Le tableau 5 montre cependant que c'est l'autre solution qu'il fallait choisir.

Appliquée à la typologie issue des poids locaux de l'analyse des indicateurs mixtes, l'analyse ne donne qu'une valeur de R inférieure à celle qu'obtient la procédure inverse : 33,22 contre 43,00. C'est pourtant cette solution de la typologie mixte améliorée par le recours aux poids locaux simples qu'il faut retenir car c'est elle qui résiste le mieux par la suite à toutes les manipulations ultérieures. En effet, comme on peut le voir sur le tableau, lorsqu'on affecte de nouveau les poids locaux simples, à la solution typologique qui donnait un R de 43,00, cette dernière retombe à la très médiocre valeur de 12,5 valeur à peine supérieure à la solution typologique sim-

²⁸ Cf. RACINE, J.B., *Op. cit.*, note 8 ; voir aussi et surtout :

JOHNSTON, R.J. (1970), Grouping and Regionalizing : Some Methodological and Technical Observations. *Economic Geography*, 46 (2, suppl.) : 293-305. *Proceedings, International Geographical Union on Quantitative Methods.*

Tableau 5
LA DÉCOUVERTE D'UNE STRUCTURE TYPOLOGIQUE OPTIMALISANT DEUX TYPOLOGIES:
SÉQUENCES DES TESTS DE CORRESPONDANCE DISCRIMINATOIRE



ple originale qui était de 9,5, et bien inférieure à la solution typologique mixte originale qui atteignait déjà la valeur intermédiaire de 16,90. En revanche la solution typologique mixte améliorée par le recours aux poids locaux simples est encore améliorée si on la nourrit à nouveau des valeurs des poids locaux mixtes, le R grim pant bien plus haut encore : à 47,5 avant les itérations discriminatoires, à 56,00 après ces dernières. Encore une fois pourquoi ne pas garder en définitive cette dernière solution qui nous fournit indiscutablement un rapport de variance intergroupe sur variance intragroupe inusité et extrêmement élevé quand on le compare aux résultats du début ? La réponse est la même : cette structure n'est pas stable : affectée des poids locaux simples elle retombe à une valeur de R bien inférieure : 21,00 soit moins que le résultat obtenu par l'amélioration de la typologie issue des indicateurs mixtes au moyen des indicateurs simples (33,22).

Avec une valeur de R tout de même considérable en regard du nombre de groupes considérés et des résultats antérieurs, cette trame typologique possède une « loi d'ordre » mesurée par son R de 33,22 qui correspond le mieux à l'ensemble des trois analyses. Face à l'évidente subjectivité des méthodes quantitatives, telle que soulignée d'abord par Johnston et dont nous avons ici une nouvelle preuve, il convient d'utiliser le résultat médian, le résultat qui tout en étant optimisé dans la plus grande mesure du possible, se conforme de plus près aux autres résultats que l'on aurait pu obtenir en suivant une autre méthode. Nous pensons qu'il s'agit là d'une règle générale et qu'en dépit de la complexité, que nous confessons volontiers, de l'analyse à laquelle nous nous sommes livrés, la leçon doit porter. Elle peut se résumer en quelques mots : une structure « optimale » ne peut être découverte directement comme le laissaient implicitement entendre dans leurs travaux ceux-là même qui ont les premiers mis au point et utilisé ces techniques en géographie, soient Emilio Casetti et Leslie J. King. Il nous semble

que le seuil qui sépare les hautes valeurs de R (de 33,22 à 56,00) des valeurs obtenues directement par les algorithmes de classification (la meilleure étant de 16,90) ne peut être franchi que par la modification d'une structure typologique par sa confrontation à d'autres attributs spatiaux de même nature issus de trames factorielles différentes. La solution que nous avons choisie pour la suite de notre travail, l'explication géographique de la typologie découverte, est donc celle qui nous paraissait la plus logique, sinon la mieux « optimisée » en termes de valeurs de R à l'une quelconque des étapes de l'analyse. Notre choix a porté sur celle qui résistait le mieux à toutes les formes de croisement et de confrontation possibles, tout en gardant en fin de compte, c'est-à-dire lorsqu'ont été testées toutes les possibilités, la plus haute somme de valeurs-propres. Le tableau 6 donne ainsi au lecteur ce que nous croyons être le meilleur découpage géographique de la zone métropolitaine de Montréal au sud du Saint-Laurent compte tenu des attributs spatiaux considérés au départ.

Rappelons cependant, pour qu'il n'y ait aucun malentendu, que le travail du géographe, du géographe explicatif, se situe bien sûr encore en aval de ce résultat, la classification n'étant évidemment qu'une étape de l'analyse scientifique. Même si nous y avons déjà introduit l'expérience et la mesure, l'explication des différents types de combinaisons spatiales se situe encore au delà. Cette explication devra se nourrir d'hypothèses issues tout à la fois de la réflexion théorique et de l'examen de la classification empirique, telle que révélée par la carte. Ces hypothèses, pour être valablement testées, devront encore pouvoir être formulées en modèles mathématiques et déductifs. Or ce n'est pas une mince affaire que de se lancer dans l'analyse typologique déductive, en dépit de la voie ouverte par le modèle que mentionne Brian J.L. Berry dans son récent article sur les principes de l'écologie factorielle. En attendant il vaudrait peut-être mieux se lancer dans la recherche d'une véritable *généralisation typologique*, qui ne soit pas liée aux seules unités d'observation de la rive sud montréalaise, mais qui, couvrant d'abord toute la couronne de la métropole, puisse être comparée à celles issues, — tant au niveau de l'analyse factorielle qu'au niveau de l'analyse typologique préconditionnée par les poids locaux des facteurs, — de la prise en considération de toutes les couronnes des grands centres urbains canadiens voire nord américains. Pour un tel travail, la découverte de véritables indicateurs géographiques simples est sans doute, du moins osons-nous l'espérer, une étape essentielle.

Tableau 6 *Poids locaux de l'analyse des indicateurs simples appliquée à la structure optimisée.*

<i>Municipalités</i>	<i>Statut socio-économique</i>	<i>Statut familial</i>	<i>Centralité économique</i>	<i>Intensité de l'occupation urbaine</i>
GROUPE I				
Brossard	0,410	0,516	0,282	-0,649
Chateauguay	-0,122	0,332	<u>0,503</u> (+)	-0,869
Chateauguay-Centre	0,030	0,515	1,050	-1,012
Saint-Hubert	0,286	0,245	0,640	-0,561
GROUPE II				
Delson	0,813	-0,002	-0,889	-0,313
Notre-Dame	1,093	0,297	-1,403	<u>-0,002</u> (-)
Sainte-Catherine	1,093	0,863	-1,352	0,402
Saint-Constant	0,850	0,310	-0,554	-0,301
Léry	0,148	-0,997	-1,033	-0,439
GROUPE III				
Boucherville	-0,132	1,394	<u>0,161</u> (-)	0,228
Candiac	-0,559	1,569	-0,445	-0,620
Saint-Bruno	-1,126	1,330	0,640	-0,485
GROUPE IV				
Greenfield Park	-0,654	-0,359	1,123	<u>0,069</u> (-)
Longueuil	-0,149	-0,917	1,135	1,843
Saint-Lambert	-1,743	-1,275	1,171	0,768
GROUPE V				
Chateauguay-Height	-0,400	-3,051	-1,021	-1,664
GROUPE VI				
Jacques-Cartier	0,964	-0,035	1,540	0,487
Lafèche	0,962	-0,447	0,036	0,028
Laprairie	0,702	-0,393	0,250	0,823
GROUPE VII				
Lemoyne	0,216	0,277	-1,338	3,523
GROUPE VIII				
Préville	-2,991	0,669	-1,778	0,346 (+)

(+) Surévalué.

(-) Sous-évalué.

RÉSUMÉ

Écologie factorielle et indicateurs géographiques

Les auteurs analysent tour à tour les contraintes algébriques et théoriques liées à l'utilisation par les géographes des algorithmes aujourd'hui traditionnels de l'écologie factorielle. Devant l'ambiguïté fondamentale de la majorité des résultats, ils en viennent à la conclusion que la règle géographique doit constamment dominer et contrôler la règle mathématique, l'outil mathématique venant pourtant féconder le contrôle critique et donner alors de meilleurs résultats géographiques. La question de leur « optimalité » est alors discutée en se référant plus particulièrement à la méthode des itérations discriminatoires. Troisième thème de réflexion enfin : celui qui conduit les auteurs à introduire la notion d'indicateurs géographiques progressivement découverts par les itérations factorielles et discriminatoires, indicateurs simples mais assortis d'un véritable arbre généalogique factoriel.

ABSTRACT

Factorial Ecology and Geographical Indicators

The authors analyse, one after the other, the algebraic and theoretical constraints of currently applied algorithms of factorial ecology as used by geographers. The continuous ambiguity of most results led them to point out that mathematical procedures should always be submitted to geographical interpretation. However the mathematical apparatus implements the critical checking and improves geographical results. The authors then proceed to the problem of their « optimality », with special reference to the method of discriminant analysis. Finally, they stress the concept of geographical indicators progressively discovered through factorial and discriminant iterations ; these indicators may seem and be simple, but they are loaded with a real factorial pedigree.